Draft Comparative Effectiveness Review

Number XX

Diagnosis and Treatment of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS)

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No.
To Be Added for Final Version

Prepared by:
To Be Added for Final Version

Investigators:
To Be Added for Final Version

AHRQ Publication No. xx-EHCxxx
<Month Year>

This report is based on research conducted by an Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. XXXXXX). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

The final report (not draft) may be used, in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

This report may periodically be assessed for the urgency to update. If an assessment is done, the resulting surveillance report describing the methodology and findings will be found on the Effective Health Care Program Web site at www.effectivehealthcare.ahrq.gov. Search on the title of the report.

Persons using assistive technology may not be able to fully access information in this report. For assistance contact EffectiveHealthCare@ahrq.hhs.gov.

None of the investigators have any affiliations or financial involvement that conflicts with the material presented in this report.

Suggested citation:
To Be Added for Final Version

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of systematic reviews to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. These reviews provide comprehensive, science-based information on common, costly medical conditions, and new health care technologies and strategies.

Systematic reviews are the building blocks underlying evidence-based practice; they focus attention on the strength and limits of evidence from research studies about the effectiveness and safety of a clinical intervention. In the context of developing recommendations for practice, systematic reviews can help clarify whether assertions about the value of the intervention are based on strong evidence from clinical studies. For more information about AHRQ EPC systematic reviews, see www.effectivehealthcare.ahrq.gov/reference/purpose.cfm

AHRQ expects that these systematic reviews will be helpful to health plans, providers, purchasers, government programs, and the health care system as a whole. Transparency and stakeholder input are essential to the Effective Health Care Program. Please visit the Web site (www.effectivehealthcare.ahrq.gov) to see draft research questions and reports or to join an e-mail list to learn about new program products and opportunities for input.

We welcome comments on this systematic review. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by email to epc@ahrq.hhs.gov.

Richard G. Kronick, Ph.D.
Director
Agency for Healthcare Research and Quality

Yen-pin Chiang, Ph.D.
Acting Deputy Director
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director
Evidence-based Practice Program
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Suchitra Iyer, Ph.D.
Task Order Officer
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Acknowledgments

The authors gratefully acknowledge the following individuals for their contributions to this project: Suchitra Iyer, PhD, Task Order Officer, Carmen Green, MD, National Institutes of Health (NIH) Working Group Chair; and the NIH Working Group.

Key Informants

This topic was nominated to the AHRQ from the NIH. Therefore, in place of Key Informants, an NIH Working Group Planning Meeting was conducted to provide input into the key questions and the scope of the report.

Technical Expert Panel

In designing the study questions and methodology at the outset of this report, the EPC consulted several technical and content experts. Broad expertise and perspectives were sought. Divergent and conflicted opinions are common and perceived as healthy scientific discourse that results in a thoughtful, relevant systematic review. Therefore, in the end, study questions, design, methodologic approaches, and/or conclusions do not necessarily represent the views of individual technical and content experts.

Technical Experts must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

The list of Technical Experts who participated in developing this report follows:
To Be Added For Final Version

Peer Reviewers

Prior to publication of the final evidence report, EPCs sought input from independent Peer Reviewers without financial conflicts of interest. However, the conclusions and synthesis of the scientific literature presented in this report does not necessarily represent the views of individual reviewers.

Peer Reviewers must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals with potential non-financial conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential non-financial conflicts of interest identified.

The list of Peer Reviewers follows:
To Be Added For Final Version

Diagnosis and Treatment of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS)

Structured Abstract

Objectives. This review summarizes current research on the clinical diagnosis of ME/CFS and the efficacy and harms of multiple medical and nonmedical interventions to treat ME/CFS in adults.

Data Source. Searches were conducted of electronic databases including MEDLINE (1988 to November 2013), PsycINFO (1988 to January 2014), the Cochrane Central Register of Controlled Trials and Cochrane Database of Systematic Reviews (through November 2013), and the Database of Abstracts of Reviews of Effects and the National Health Sciences Economic Evaluation Database (through the 4th quarter 2013). The searches were supplemented by reviewing reference lists and requesting scientific information from drug and device manufacturers.

Review Methods. Abstracts and full-texts were reviewed by two investigators for inclusion based on predefined criteria. Discrepancies were resolved through discussion and consensus, with a third investigator making the final decision as needed.

Results. 5,902 potentially relevant articles were identified, 914 selected for full-text review, and 64 studies (in 71 publications) included, 28 observational studies on diagnosis and 36 trials on interventions. Multiple case definitions have been used to define ME/CFS and those that require the symptoms of post-exertional malaise and neurological and autonomic manifestations appear to represent a more severe subset of the broader ME/CFS population. The artificial neural network test had good ability to discriminate between patients with ME/CFS compared with those without the condition (sensitivity 0.95; specificity 0.85 and accuracy 0.90), but no test has been adequately evaluated in a large population with diagnostic uncertainty to determine validity and generalizability. A diagnosis of ME/CFS is associated with broad psychosocial consequences. Of the 36 trials on interventions, rintatolimod improved measures of exercise performance, compared with placebo; cognitive and behavioral therapy (CBT) and graded exercise treatment (GET) compared with no treatment, relaxation or support were found to improve fatigue, function, and quality of life, while CBT also improved employment outcomes. Other interventions either provided no benefit or evidence was insufficient to draw conclusions. Although adverse effects were not well reported across trials, GET compared with CBT or control groups was associated with a higher number of reported adverse events and withdrawal rates in several trials.

Limitations. Diagnostic tests were not well studied in a broad spectrum of patients. Intervention studies were scarce and most were either fair- or poor-quality and measured outcomes using heterogeneous methods making it difficult to compare results across studies.

Conclusions. No current diagnostic tool or method has been adequately tested to identify patients with ME/CFS when diagnostic uncertainty exists. CBT and GET have shown some benefit whereas other interventions have insufficient evidence to guide clinical practice. GET appears to be associated with harms in some patients whereas the negative effects of being given a diagnosis of ME/CFS appear to be more universal.

Contents

Tables

Figures

Appendixes

Executive Summary

Background

Myalgic encephalomyelitis (ME) and/or chronic fatigue syndrome (CFS) is a condition characterized by chronic and disabling fatigue as well as various additional manifestations including pain, sleep disturbance, neurological and cognitive changes, motor impairment, and altered immune and autonomic responses.[1-3] Experts consider post-exertional malaise (PEM) and memory or concentration problems critical components.[4]

The term ME was first used in the 1930s after an outbreak of neuromyesthenia and CFS was first coined in the 1980s.[5-7] Attempts to describe the condition based on possible underlying etiologies led to additional terms including post viral fatigue syndrome and chronic fatigue immune dysfunction syndrome.[1,3,5,6] The most recent international consensus report advocates moving away from the term CFS in favor of ME to better reflect an underlying pathophysiology involving widespread inflammation and neuropathology, and to embrace the two terms as synonymous.[2] However, others believe that ME is a subset of CFS and represents a more severe form of the same disease.[4] Some feel that the lack of specificity surrounding the name, CFS, may delegitimize and negatively characterize the condition, and stigmatize patients.[8] For this review, ME and CFS will be used synonymously (ME/CFS) and will include the populations(s) studied under either of these terms, recognizing that issues regarding terminology are currently unresolved.

Uncertainty persists regarding the etiology and whether the condition reflects a single pathologically discrete syndrome, subsets of the same illness, or a nonspecific condition shared by other disease entities. Some suggest that an inciting event triggers an immune response and promotes immune and/or neuroendocrine dysregulation that perpetuates the body's response and symptom experience that becomes ME/CFS.[9,10] Viral etiologies have been predominantly studied based on the observation that the majority of patients report a sudden onset of symptoms associated with a preceding febrile illness and enlarged lymph nodes. However, no specific virus or other infectious agent has been identified, and not all patients experience a preceding febrile illness.[9] Numerous studies have attempted to identify risk factors for developing ME/CFS but a systematic review in 2008 of 11 studies that assess multiple predictors found no evidence of any definitive factors.[11] This review is not intended to address the question of etiology nor underlying factors that lead to the onset or perpetuation of ME/CFS but rather to focus on the diagnosis and treatment of this syndrome.

Currently diagnosing a patient with ME/CFS relies on the use of a set of clinical criteria (case definitions) to distinguish ME/CFS from other conditions that may also present with fatigue. There are currently eight published case definitions that have evolved since first one was published by the Centers for Disease Control and Prevention (CDC) in 1988.[6] All include persistent fatigue not attributable to a known underlying medical condition, as well as additional clinical signs and symptoms that do not all need to be present to establish the diagnosis.[4] As with other medical syndromes that involve a multitude of symptoms and lack a definitive diagnostic test, differentiating one disease state from another similar or overlapping condition becomes a challenge. Some clinicians are reluctant to diagnose ME/CFS believing that the diagnosis will harm the patient or that the patient be inappropriately labeled.[12] This makes the prevalence of ME/CFS difficult to assess.[13,14] The CDC reported a U.S. prevalence rate of 0.3 percent corresponding to over 1,000,000 adults.[15] By using different case definitions, the rate may be as high as 2.5 percent.[5,16] A recent systematic review found that when using the same case

definition (CDC Fukuda, 1994), the prevalence was higher when determined by self report (3.28%; 95% confidence interval [CI] 2.24 to 4.33) compared with clinical assessment (0.76%; 95% CI, 0.23 to 1.29).[17] ME/CFS is more common among women with the average age of diagnosis between 30 and 40 years.[13] Childhood ME/CFS is uncommon and although symptoms may be similar, prognosis appears to be different.[18,19] Although the natural history is not well studied, symptoms and disability in adults tend to persist over time,[20] with approximately 5 percent (0-31%) of adult patients fully recovering despite 40 percent (8-63%) of adult patients improving over time[21] in contrast to childhood studies that suggest that over 50 percent of patients will recover within 6 months.[18] Economic impact is considerable with most adult patients never returning to work.[9,21]

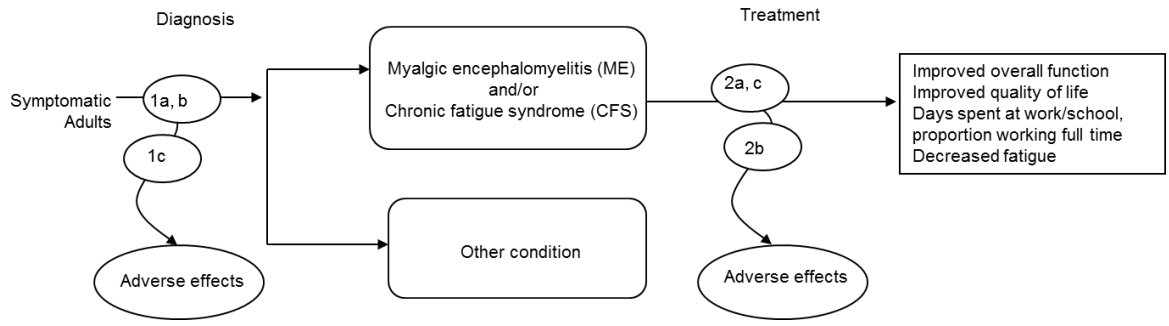
Currently there are no medications for the treatment of ME/CFS approved by the U.S. Food and Drug Administration (FDA), but many have been used 'off-label' (without review and approval), and some have been obtained from other countries and are not currently approved for any indication in the United States (i.e., Isoprinosine, rintatolimod). In a survey by the FDA, patients identified treatments that fell into two broad categories, those intended to treat the underlying cause of the disease and those targeting specific symptoms.[22] Medications to treat underlying causes include immune modulators such as rituximab and the investigational drug rintatolimod, and antiviral and antibiotic medications. Interventions targeting symptoms include medications to treat pain, fatigue, autonomic dysfunction, and sleep dysfunction, and non-drug therapies such as yoga, exercise techniques, counseling on pacing strategies, and mental exercises.[22] In practice, there are wide variations in the clinical management of patients, and many patients receive a multifaceted approach to treatment.

The variable symptomatology of ME/CFS, lack of a clearly identifiable etiology and disease process, and no accepted diagnostic tests or treatments have challenged researchers and clinicians in their attempts to better understand the condition and treat patients. This systematic review was commissioned by the Office of Disease Prevention (ODP) at the National Institutes of Health (NIH) to inform a Pathways to PreventionWorkshop on ME/CFS. This review summarizes the research on diagnosis and treatment of ME/CFS including the methods and criteria used to diagnose ME/CFS, their utility in clinical practice, the harms associated with carrying a diagnosis of MECFS, and the evidence on treatment effectiveness and associated harms. It identifies areas of future research needed to better inform the diagnostic process and treatment strategies. This report is not intended to be used or likely to be useful to develop criteria for disability or insurance

Scope of Review and Key Questions

This topic was nominated for review by the NIH and focuses on diagnosis and treatment for ME/CFS. The analytic framework (**Figure A**) and key questions used to guide this review are shown below. The analytic framework shows the target populations, interventions, and health outcomes we examined, with numbers corresponding to the key questions.

Figure A. Analytic framework



The following key questions are the focus of the report:
Key Question 1. What methods are available to clinicians to diagnose ME/CFS and how do the use of these methods vary by patient subgroups?

Key Question1a. What are widely accepted diagnostic methods and what conditions are required to be ruled out or excluded before assigning a diagnosis of ME/CFS ?

Key Question 1b.What is the accuracy and concordance of diagnostic methods?

Key Question 1c. What harms are associated with diagnosing ME/CFS?

Key Question 2a. What are the (a) benefits and (b) harms of therapeutic interventions for patients with ME/CFS and how do they vary by patient subgroups?

Key Question 2c. What are the characteristics of responders and non-responders to interventions?

Methods

This systematic review follows the methods of the Agency for Healthcare Research and Quality (AHRQ) Methods Guide for Effectiveness and Comparative Effectiveness Reviews (hereafter "AHRQ Methods Guide").[23]

Topic Development and Refinement

The initial key questions were developed in conjunction with the NIH and AHRQ, and revised with input from a Technical Expert Panel (TEP) convened for this report. The TEP consisted of experts in six disciplines and two patients who disclosed no conflicts of interest that precluded participation. The TEP did not contribute to reviewing the evidence of writing of the report.

With input from the TEP, the NIH, and AHRQ, the final protocol was developed and posted on the AHRQ web site on May 1, 2014 at: http://effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?productid=1906&pageaction=displayproduct. The protocol was also registered in the PROSPERO international database of prospectively registered systematic reviews.[24]

Literature Search Strategy

A research librarian conducted searches in Ovid MEDLINE (1988 to November 2013), PsycINFO (1988 to January 2014), the Cochrane Central Register of Controlled Trials and

Cochrane Database of Systematic Reviews (through November 2013), and the Database of Abstracts of Reviews of Effects and the National Health Sciences Economic Evaluation Database (through the 4th quarter 2013). Searches were supplemented with hand-searches of reference lists of relevant studies. In addition, scientific information packets were requested from drug and device manufacturer who potentially had data on the use of medications or devices for ME or CFS, who had the opportunity to submit data using the portal for submitting scientific information packets on the Effective Health Care Program Web site. Seventeen submissions were received. Library searches will be updated while the draft report is under external review.

Process for Study Selection

Criteria for inclusion and exclusion of studies was developed based on the key questions and the populations, interventions, comparators, outcomes, timing, types of studies, and setting (PICOTS) approach. Papers were selected for review if they were about diagnosis or treatment of ME or CFS in adult populations, were relevant to a key question, and met the prespecified inclusion criteria. Studies of nonhuman subjects and studies with no original data were excluded. Abstracts were reviewed by two investigators for inclusion for each key question. Full-text articles were obtained for all studies that either investigator identified as potentially meeting inclusion criteria. Two investigators independently reviewed all full-text articles for final inclusion. Inclusion was restricted to English language articles. Discrepancies were resolved through discussion and consensus, with a third investigator making the final decision if necessary.

Population and Conditions of Interest

For Key Question 1, symptomatic adults, ages 18 years or older, with fatigue were included. For Key Question 2, adults, ages 18 years or older, diagnosed with ME, CFS, or both by fulfilling criteria from at least one of the case definitions and without another underlying diagnosis were included.

Interventions and Comparisons/Study Designs

For Key Question 1, diagnostic accuracy or concordance studies comparing case definitions (e.g., Fukuda/CDC, Canadian, International, and others) with one another were included.

For Key Question 2, randomized trials comparing medication management (immune modulators, beta blockers, antidepressants, anxiolytics, stimulants, other), counseling and behavior therapy, graded exercise programs, and complementary and alternative medicine (CAM) approaches (acupuncture, relaxation, massage, other) and with placebo, no treatment, usual care, or other active interventions, including combination therapies and head-to-head trials were included. For harms, cohort studies with control groups were also included.

Outcomes

For Key Question 1, outcomes of diagnostic accuracy or concordance include sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio, negative likelihood ratio, C-statistic, receiver operator curve (ROC) and area under curve (AUC), net reclassification index, concordance, and any potential harm from diagnosis (such as psychological harms, labeling, risk from diagnostic test, misdiagnosis).

For Key Question 2, outcomes that are patient-centered and include overall function (i.e., 36-item Short Form Survey [SF-36]), quality of life, days spent at work or school, proportion

working full- or part-time, and fatigue (Multidimensional Fatigue Inventory 20-item [MFI-20] or similar). Individual symptom-based outcomes were excluded. Included harms are withdrawals, withdrawals due to adverse events, and rates of adverse events due to interventions.

Timing

Intervention studies, for Key Question 2, must have a minimum duration of 12 weeks of treatment to be included given the cyclical nature of the condition. There was no duration or timing restriction on studies included for Key Question 1.

Setting

Studies for all Key Questions had to be conducted in a clinical setting or a setting that was generalizable to clinical practice settings. Studies conducted with in-patients or institutionalized individuals were excluded.

Data Extraction and Data Management

The following information was extracted from included studies into evidence tables: study design, setting, inclusion and exclusion criteria, population characteristics (including sex, age, race, and co-morbidities), sample size, duration of followup, attrition, intervention characteristics, case definition used for diagnosis, duration of illness, and results. Data extraction for each study was performed by two investigators: the first investigator extracted the data, and the second investigator independently reviewed the extracted data for accuracy and completeness.

For studies of diagnostic accuracy and concordance, when reported, we extracted relative measures of risk (relative risk [RR], odds ratio [OR], hazards ratio [HR]), ROC, and AUC. The area under the receiver operating characteristic (AUROC), which is based on sensitivities and specificities across a range of test results, is a measure of discrimination, or the ability of a test to distinguish people with a condition from people without the condition.[25,26] An AUROC of 1.0 indicates perfect discrimination, and an AUROC of 0.5 indicates complete lack of discrimination. Interpretation of AUROC values between 0.5 and 1.0 is somewhat arbitrary, but a value of 0.90 to 1.0 has been classified as excellent, 0.80 to <0.90 as good, 0.70 to <0.80 as fair, and <0.70 as poor.[27]

Individual Study Quality Assessment

The quality of each study was assessed based on predefined criteria adapted from methods proposed by the U.S. Preventive Services Task Force. The criteria used are consistent with the approach recommended by AHRQ in the AHRQ Methods Guide.[23] The term "quality" was used rather than the alternate term "risk of bias;" both refer to internal validity. Two investigators independently assessed the quality of each study. Discrepancies were resolved through discussion and consensus, with a third investigator making the final decision if necessary.

To determine the quality of each study evaluating diagnostic tests, we used questions from the AHRQ Methods Guide for Medical Test Reviews and adapted them to improve their clinical relevance to ME/CFS.[28] Quality was based on whether it evaluated a representative spectrum of patients, whether it enrolled a random or consecutive sample of patients meeting prespecified criteria, whether it used a credible reference standard, whether the same reference standard was applied to all patients, whether the reference standard was interpreted independently from the test

under evaluation, and whether thresholds were prespecified.[23,29,30] Descriptive papers that compared diagnostic criteria and reported harms were not quality rated.

The quality of intervention trials was based on the methods used for randomization, allocation concealment, and blinding; the similarity of compared groups at baseline; maintenance of comparable groups; adequate reporting of dropouts, attrition, crossover, adherence, and contamination; loss to followup; the use of intent-to-treat analysis; and ascertainment of outcomes.[23,30]

Following assessment of individual quality criteria, individual studies were rated as "good," "fair," or "poor" quality, as defined below.[23,28]

Good-quality studies are considered likely to be valid. Good-quality studies clearly describe the population, setting, interventions, and comparison groups; use a valid method for allocation of patients to interventions; clearly report dropouts and have low dropout rates; use appropriate methods for preventing bias; assess outcomes blinded to intervention status; and appropriately measure outcomes and fully report results.

Fair-quality studies have some methodological deficiencies, but no flaw or combination of flaws judged likely to cause major bias. The study may be missing information, making it difficult to assess its methods or assess limitations and potential problems. The fair-quality category is broad, and studies with this rating vary in their strengths and weaknesses: the results of some fair-quality studies are likely to be valid, while others are probably *in*valid.

Poor-quality studies have significant flaws that may invalidate the results. They have a serious or "fatal" flaw in design, analysis, or reporting; large amounts of missing information; or discrepancies in reporting. The results of these studies are judged to be at least as likely to reflect flaws in the study design as true effects of the interventions under investigation. Poor-quality studies were not excluded *a priori*, but they were considered to be the least reliable studies when synthesizing the evidence, particularly when discrepancies between studies were present.

Assessing Research Applicability

Applicability is defined as the extent to which the effects observed in published studies are likely to reflect the expected results when a specific intervention is applied to the population of interest under "real-world" conditions.[23] It is an indicator of the extent to which research included in a review might be useful for informing clinical and/or policy decisions in specific situations. Applicability depends on the particular question and the needs of the user of the review. There is no generally accepted universal rating system for applicability. In addition, applicability depends in part on context. Therefore, a rating of applicability (such as "high" or "low") was not assigned because applicability may differ based on the user of this review. Rather, factors important for understanding the applicability of studies were recorded, such as how similar patients were to the population of interest, how large the sample size was, and the characteristics of the clinical setting.[31] The funding source for treatment trials was also recorded.

Data Synthesis

Results of diagnostic accuracy studies (such as creating summary AUROCs) were not quantitatively pooled due to differences in methods, case definitions, and heterogeneity in the outcomes. Instead, descriptive statistics were used, such as the median sensitivity and specificity at specific cutoffs and reported AUROCs, along with associated ranges, and calculated positive and negative likelihood ratios based on the median sensitivities and specificities. For the results of intervention trials, the appropriateness of meta-analysis was determined by considering the

internal validity of the studies and the heterogeneity among studies in design, patient population, interventions, and outcomes. Appropriate measures were chosen based on the type of data for meta-analysis, according to the guidance for the EPC program.[32] Random-effects models were used to estimate pooled effects, except when only two studies were available we chose to not pool the results.[33,34] We calculated pooled relative risks where the data were reported as proportions of dichotomous outcomes (e.g., proportion with improvement in intervention and control groups). For continuous outcomes, we calculated pooled weighted mean differences using the means and standard deviations (e.g., mean change in function based on a scale). The Q statistic and the I-squared statistic (the proportion of variation in study estimates due to heterogeneity) were calculated to assess heterogeneity in effects between studies.[35] When statistical heterogeneity was found, we explored the reasons by using subgroup analysis. In meta-analysis, we combined relative risks and odds ratios for such outcomes.

Grading the Body of Evidence for Each Key Question

The overall strength of evidence was assessed for key question and outcome in accordance with the AHRQ Methods Guide.[23,28] Strength of evidence was based on the overall quality of each body of evidence, the study limitations (graded low, moderate, or high); the consistency of results between studies (graded consistent, inconsistent, or consistency unknown when only one study was available); the directness of the evidence linking the intervention and health outcomes (graded direct or indirect); the precision of the estimate of effect, based on the number and size of studies and confidence intervals for the estimates (graded precise or imprecise); and the whether reporting bias was suspected (graded suspected or undetected). There was no way to formally assess for publication bias due to small number of studies, methodological shortcomings, or differences across studies in designs, measured outcomes, and other factors. Studies included to answer Key Question 1 were not formally evaluated for the strength of evidence, but key concepts of strength of evidence are discussed.

The strength of evidence was rated for Key Questions 2 and 2a using the four categories recommended in the AHRQ Methods Guide:[23,28] A "high" grade indicates high confidence that the estimate of effect lies close to the true effect for this outcome. The body of evidence has few or no deficiencies and the findings are stable (i.e., another study would not change the conclusions). A "moderate" grade indicates moderate confidence that the estimate of effect lies close to the true effect for this outcome. The body of evidence has some deficiencies and findings are likely to be stable, but some doubt remains. A "low" grade indicates low confidence that the estimate of effect lies close to the true effect for this outcome. The body of evidence has major or numerous deficiencies (or both) and additional evidence is needed before concluding either that the findings are stable or that the estimate of effect is close to the true effect. An "insufficient" grade indicates inability to estimate an effect, or no confidence in the estimate of effect for this outcome. No evidence is available or the body of evidence has unacceptable deficiencies, precluding reaching a conclusion.

Peer Review and Public Commentary

Experts in ME/CFS, individuals representing important stakeholder groups, and TEP members have been invited to provide external peer review of this systematic review. The AHRQ Task Order Officer and a designated Evidence-based Practice Center Associate Editor will also provide comments and editorial review. To obtain public comment, the draft report will be posted on the AHRQ web site for 4 weeks. A disposition of comments report detailing the

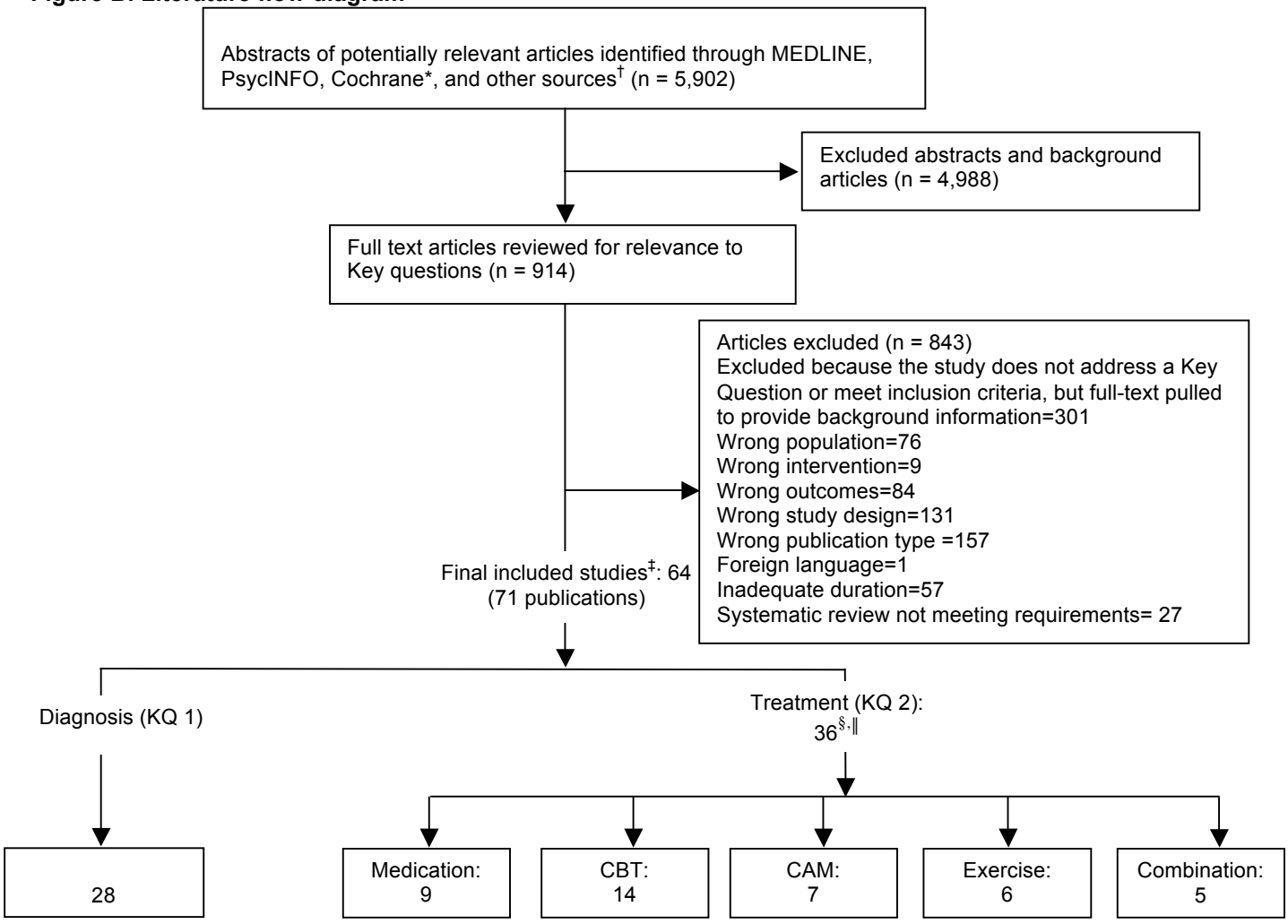
authors' responses to the peer and public review comments will be made available after AHRQ posts the final systematic review on the public web site.

Results

Results of Literature Searches

Results of the literature search and selection process are summarized in the literature flow diagram (**Figure B**). Database searches results in 5,902 potentially relevant articles. After dual review of abstracts and titles, 914 articles were selected for full-text review. After dual review of full text articles, 64 studies (in 71 publications) met inclusion criteria and were included. Data extraction and quality assessment tables for included studies by key question are available in **Appendixes G** and **H**.

Figure B. Literature flow diagram



CAM = complementary alternative medicine; CBT = cognitive behavioral therapy; KQ = key question.

*Cochrane databases include the Cochrane Central Register of Controlled Trials, Database of Abstracts of Reviews of Effects, Health Technology Assessment, National Health Sciences Economic Evaluation Database, and the Cochrane Database of Systematic Reviews.
†Identified from reference lists, hand searching, suggested by experts, etc.
‡Studies that provided data and contributed to the body of evidence were considered 'included'.
§ Studies may have more than one published article, this number indicates the number of unique studies included; there were a total of 43 articles included.
‖Studies may have provided data for more than one treatment area.

Description of Included Studies

Of the 64 studies included in this review, 28 observational studies addressed Key Question 1 pertaining to aspects of diagnosis. Most were of fair-quality, enrolled predominantly female patients, had small sample sizes, and were conducted in the United States and Western Europe. Thirty-six randomized trials were included for Key Question 2, addressing the benefits and harms of interventions to treat ME/CFS (9 for medications, 14 for counseling or behavioral therapies, 7 for CAM, 6 for exercise, and 5 comparing therapies). Most were of fair- or poor-quality, enrolled predominantly female patients from ME/CFS specialty clinics based on the CDC (Fukuda, 1994) case definition, had small sample sizes, and were conducted in the United States and Western Europe.

Key Question 1

What methods are available to clinicians to diagnose ME/CFS and how do the use of these methods vary by patient subgroups?

- Eleven cross-sectional and longitudinal studies with comparison groups evaluated methods currently used to diagnose ME/CFS and provided data on discriminative value (ROC/AUC), sensitivity/specificity, or concordance of diagnoses. Most were of fair quality, small sample size, and did not include patients with diagnostic uncertainly. No studies evaluated a diagnostic test for ME/CFS using an adequate size and spectrum of patients and no studies demonstrated an accurate and reliable method for identifying patients or subgroup of patients with ME/CFS.
- The artificial neural network test using a combination of 24 symptoms had good sensitivity, specificity, and accuracy in diagnosing ME/CFS based on one good-quality study derived and validated in a small but broad-spectrum of patients. ME/CFS symptoms with greatest single-item accuracy were acute onset of fatigue and sore throat. Validation in different populations would improve the confidence that these 24 symptoms in the artificial neural network can discriminate between patients with ME/CFS and those without the condition when diagnostic uncertainty exists.
- The Schedule of Fatigue and Anergia for CFS scale (SOFA-CFS) and certain subscales of the SF-36 may be promising as tools for systematic evaluation of certain components of the ME/CFS diagnosis. However, these require further testing in broader populations, and integration with the full set of clinical criteria that are used for ME/CFS diagnosis.
- Specific SF-36 subscales may correlate with symptoms of post-exertional malaise (PEM), based on one small fair-quality study, but case-control design precludes any valid conclusion about the use of SF-36 subscales to identify which ME/CFS patients will fail to recover at 1-day or 1-week after cardio-pulmonary exercise testing (CPET). Other self-reported symptom scales did not add to the diagnostic strategy given methodological

limitations including case-control study designs, small sample size, and unclear reporting of study details.

- Serum biomarkers (including cortisol response to dexamethasone suppression testing, plasma and salivary cortisol response to insulin, insulin tolerance testing, plasma adrenocorticotropin hormone level, pro-inflammatory cytokine response to psychological stress, and RN'ase L-isoform) were studied in four small, single studies in populations without diagnostic uncertainty, providing insufficient evidence to determine their diagnostic usefulness in ME/CFS.

Key Question 1a

What are widely accepted diagnostic methods and what conditions are required to be ruled out or excluded before assigning a diagnosis of ME/CFS?

- Eight case definitions of ME/CFS that include a set of clinical criteria are used to identify patients with ME/CFS and used by clinicians to distinguish ME/CFS from other conditions that also present with fatigue.
- Despite the fact that most ME/CFS case definitions require that other conditions be excluded prior to assigning a diagnosis of ME/CFS, no studies compared strategies for ruling out alternative diagnoses.

Key Question 1b

What is the accuracy and concordance of diagnostic methods?

- Diagnostic methods were evaluated in nine studies, primarily observational cohort and descriptive in nature. Six studies found that the symptoms reported by different case definitions varied. Populations identified by criteria labeled as ME[2] or ME/CFS[1,7,,9] had more severe symptoms or more functional impairment than those labeled as CFS[3,5,,6,8,] criteria. There is no diagnostic gold standard for ME/CFS and no studies evaluated the accuracy of current diagnostic methods.
- Three studies that compared CFS patients with healthy controls also found differences in symptom reporting using various self-reported symptom scales. These results suggest that the Fatigue Impact Scale (FIS), Chalder Fatigue Scale, Hospital Anxiety and Depression scale (HADS) -depression subscale (HADS-D), and certain SF-36 subscales or combinations of SF-36 variables with Zung Depression Scale might be reasonable candidates for further evaluation as diagnostic tests but will need to be evaluated in a broad spectrum of patients with diagnostic uncertainty to determine their ability to differentiate between conditions. No studies evaluated if different diagnostic methods could adequately identify clinical subgroups of patients. [36-38]

Key Question 1c

What harms are associated with diagnosing ME/CFS?

- Five studies consistently found that patients with CFS feel stigmatized by their diagnosis in terms of financial stability, work opportunities, perceived judgments on their character, social isolation, and interactions with the health care system.
- Prejudice and stereotypes within the medical profession have been identified in two studies; medical trainees and mental health practitioners make judgments about a patient's condition based on the name it carries (ME, CFS, or other) and what treatment is being given. There is a substantial burden of misdiagnosis among the ME/CFS population.

Key Question 2

What are the (a) benefits and (b) harms of therapeutic interventions for patients with ME/CFS and how do they vary by patient subgroups?

Thirty-six trials of interventions for patients with ME/CFS in 43 publications met inclusion criteria; nine of medications, 14 of counseling and behavior therapies, seven of CAM interventions, six of exercise programs, and five of head-to-head or combinations of these interventions. Seven were rated good-quality, while 24 were rated fair- and five poor-quality. Trials enrolled from 22 to 641 patients with ME/CFS and most (26/36, 72%) used the CDC (Fukuda, 1994) case definitions diagnostic criteria. Outcomes measured included the SF-36 physical functioning subscale, Medical Outcome Study Short Form (MOS-SF), Checklist of Individual Strength (CIS), Profile of Mood States (POMS) fatigue subscale, Karnofsky Performance Scale (KPS), and Sickness Impact Profile 8-item (SIP-8) scale to measure overall function; MFI-20, Chalder Fatigue Scale, Krupp Fatigue Severity Scale (FSS), Fatigue Impact Scale (FIS), and Visual Analogue Scale (VAS) to measure fatigue; Quality of Life Inventory (QOLI), Quality of Life Index (QLI), Quality of Life Scale (QLS), EuroQol Scale, Global Wellness Scale, Short Form 12-item Health Survey (SF-12), and Fibromyalgia Impact Questionnaire (FIQ) to measure quality of life; Clinical Global Impression Change (CGI) scales to measure improvement over time; and the Work and Social Adjustment Scale to measure impairment in work

Nine randomized trials met inclusion criteria for medical treatment of ME/CFS, although none of the medications have been approved by the FDA for this indication. A small trial of valganciclovir enrolled participants with suspected viral onset of ME/CFS and elevated antibody titers and reported improved fatigue compared with placebo based on one scale, but no differences for other measures. Two trials of rintatolimod, an immune modulator, enrolling severely debilitated participants found improvement in measures of exercise, performance, and activities of daily living, and reduction of other medications for relief of CFS symptoms, although some of these comparisons were of borderline statistical significance, other measures did not differ, and trials were small. Small single trials of Isoprinosine, galantamine, hydrocortisone, immunoglobulin G, and fluoxetine did not show significant improvement compared with placebo. Harms of medications included suppression of adrenal glucocorticoid responsiveness, increased appetite, weight gain, and difficulty sleeping with hydrocortisone; flu-

like syndrome, chills, vasodilatation, dyspnea, and dry skin with rintatolimod; headaches with immunoglobulin G; and withdrawals due to adverse events with fluoxetine. The strength of evidence for rintatolimod is low for the measures of exercise performance based on two small trials. Strength of evidence for other medications is insufficient because each medication was evaluated by only one small trial with important methodological limitations and few differences were found between treatment and placebo groups.

Based on 13 trials, cognitive and behavioral therapy (CBT), either group or individual; self-instruction booklets; pragmatic rehabilitation: peer-to-peer counseling: and symptom consultation provide improvement in fatigue, function, quality of life, and employment in adult patients with ME/CFS. When combining all studies comparing any type of counseling to no treatment, support, relaxation, or adaptive pacing there is moderate strength of evidence that counseling improves fatigue (8/15 trials showed positive effect), and global improvement and global improvement (3/3 trials showed positive effect), and low strength evidence that counseling improves functioning measures (6/13 trials showed positive effect; weighted mean difference of 7.73 (95% CI, 3.58 to 11.87), quality of life (2/5 trials showed positive effect; 1/5 showed mixed results on measures), and employment outcomes (positive benefit: 2/2 work impairment; 2/3 hours worked; 0/2 proportion working full- or part-time at 1 or 5 years). Harms of counseling and behavioral therapies were poorly reported but there is low strength of evidence that CBT is not associated with harms based on one moderate-sized trial.

The evidence on CAM interventions is insufficient to draw conclusions as all interventions included only single trials, all were of small sample size, and most had significant methodological limitations. Two fair-quality trials, one with homeopathy and one with L-carnitine preparations, found improvement in some measure of fatigue and/or function, with no differences found in other measures. One good-quality trial found that improvement in function was less in those patients aware that they were not receiving distant healing. All other trials of CAM interventions found no significant improvements compared with placebo, usual care, or an alternative CAM approach. Adherence was low in one trial of a low sugar/low yeast diet but otherwise adherence and harms were not well reported.

Graded exercise treatment (GET) was superior to no exercise or relaxation/flexibility or adaptive pacing in measures of function (moderate strength), global improvement (low strength), and fatigue (low strength) based on one good-quality and three fair-quality randomized trials. While qigong exercise compared with no qigong found improvement in some measures of fatigue and home orthostatic training improved physiological measures, neither showed improvement in function; given that this is based on single small trials, it represents an insufficient strength of evidence.

There is low strength of evidence that GET and CBT or cognitive therapy had similar results on measures of fatigue and function in one good-quality and two fair-quality head-to-head trials with evidence of superiority over usual care or adaptive pacing in two of the three trials. GET was superior to fluoxetine on measures of fatigue and function in one fair-quality trial but represents an insufficient strength of evidence given that it was a single study of small sample size. Harms were not well reported leaving insufficient evidence on the harms of GET although patients receiving GET reported more adverse effects compared with CBT, adaptive pacing, or usual care in one good-quality trial and there were more withdrawals in the GET group in several trials.

Key Question 2c

What are the characteristics of responders and non-responders to interventions?

One small fair-quality trial found that those who had lower functional impairment, less fatigue, and less pain at baseline were more likely to improve after group CBT.

Table A. Summary of evidence

Key question Outcome	Study design Number of studies (n)*	Findings and direction of effect	Strength of evidence grade
KQ1. What methods are available to clinicians to diagnose ME/CFS and how do the use of these methods vary by patient subgroups?			
Available methods	11 cross-sectional and case control studies (n=1,470)	1 good-quality study conducted with a spectrum of patients that included conditions similar to ME/CFS found that the artificial neural network test using a combination of 24 symptoms had good sensitivity, specificity, and accuracy. In that study, CFS symptoms with greatest single-item accuracy were acute onset of fatigue and sore throat.	Not applicable
Available methods	11 cross-sectional and case control studies (n=1,470)	The SOFA-CFS and certain subscales of the SF-36 may be promising for identification of certain components of the CFS criteria, but require further testing in broader populations.	Not applicable
KQ1a. What are widely accepted diagnostic methods and what conditions are required to be ruled out or excluded before assigning a diagnosis of ME/CFS ?	No studies	No studies evaluated strategies for working up a patient.	Not applicable
KQ1b. What is the accuracy and concordance of diagnostic methods?	9 observational/ descriptive studies (n=1,178)	6 studies found that the symptoms reported by different case definitions varied. In general, populations defined by ME or ME/CFS criteria had more severe symptoms or more functional impairment than those defined by CFS criteria alone.	Not applicable
KQ1c. What harms are associated with diagnosing ME/CFS?			
Psychological harm, including stigma from label	5 observational studies (n=677)	5 studies found that patients with CFS feel stigmatized by their diagnosis in terms of financial stability (1), work opportunities (1), perceived judgments on their character (1), social isolation (2), or interactions with the health care system (3).	Not applicable
Misdiagnosis	1 observational study (n=68)	1 study identified a substantial burden of misdiagnosis among the CFS population.	Not applicable
Risk from diagnostic test	No studies	No studies identified that reported objective risks directly related to the process of conducting a diagnostic test for CFS.	Not applicable

Key question Outcome	Study design Number of studies (n)*	Findings and direction of effect	Strength of evidence grade
Prejudice and stereotyping	2 observational studies (n=246)	2 studies identified prejudice and stereotypes within the medical profession; medical trainees and mental health practitioners make judgments about a patient's condition based on the name it carries (ME, CFS, or other) and what treatment is being given.	Not applicable
KQ2a. What are the benefits of therapeutic interventions for patients with ME/CFS and how do they vary by patient subgroups?			
Galantamine vs. placebo			
Decreased fatigue and improved quality of life	1 RCT (n=423)	No significant differences between 4 intervention groups and placebo.	Insufficient
Global improvement	1 RCT (n=423)	No significant differences between 4 intervention groups and placebo.	Insufficient
Improved overall function, increased days spent at work/school and proportion working full- or part-time	No studies	No studies.	Insufficient
Hydrocortisone vs. placebo			
Improved overall function, decreased fatigue, and improved quality of life	1 RCT (n=68)	No significant differences between intervention and placebo.	Insufficient
Increased days spent at work/school and proportion working full- or part-time	No studies	No studies.	Insufficient
Hydrocortisone + fludrocortisone vs. placebo			
Improved overall function, decreased fatigue, and improved quality of life	1 RCT (n=80)	No significant differences between intervention and placebo.	Insufficient
Increased days spent at work/school and proportion working full- or part-time	No studies	No studies.	Insufficient
Immunoglobulin G vs. placebo			
Improved overall function	1 RCT (n=28)	Significantly better scores on SF-36 social functioning scale after intervention compared with placebo ($p<0.05$), but no difference on physical functioning scale.	Insufficient

Key question Outcome	Study design Number of studies (n)*	Findings and direction of effect	Strength of evidence grade
Improved fatigue and quality of life, increased days spent at work/school and proportion working full- or part-time	No studies	No studies.	Insufficient
Rintatolimod vs. placebo			
Improved overall function	1 RCT (n=84)	Significant increase in activities of daily living after intervention compared with placebo (23% vs. 14%, p=0.034), but no difference in change in KPS scores from baseline.	Insufficient
Increased exercise work capacity	2 RCT (n=316)	The intervention group compared with placebo had significant increases in exercise duration (10% vs. 2%, p=0.007), exercise work (12% vs. 6%, p=0.011), and cardiopulmonary exercise tolerance (37% vs. 15%, p=0.047).	Low
Improved quality of life, increased days spent at work/school and proportion working full-or part-time	No studies	No studies.	Insufficient
Valganciclovir vs. placebo			
Decreased fatigue	1 RCT (n=30)	Significant decrease in fatigue based on FSS scores decreasing in intervention group compared with placebo (mean change from baseline: -0.06 vs. 0.02, p=0.006).	Insufficient
Improved overall function	1 RCT (n=30)	No significant differences between intervention and placebo.	Insufficient
Improved quality of life, increased days spent at work/school and proportion working full-or part-time	No studies	No studies.	Insufficient
Isoprinosine vs. placebo			
Improved overall function and decreased fatigue	1 RCT (n=15)	No significant differences between intervention and placebo.	Insufficient
Improved quality of life, increased days spent at work/school and proportion working full-or part-time	No studies	No studies.	Insufficient
Fluoxetine vs. placebo			
Improved overall function	1 RCT (n=68)	No significant differences between intervention and placebo.	Insufficient
Decreased fatigue	1 RCT (n=68)	No significant differences between intervention and placebo.	Insufficient

Key question Outcome	Study design Number of studies (n)*	Findings and direction of effect	Strength of evidence grade
CBT/counseling vs. no treatment or support or relaxation or adaptive pacing			
Improved overall function	12 RCT (n=1,637)	Results were mainly positive, but mixed. When 8 trials using the SF-36 physical functioning subscale were pooled there was a significant effect for the intervention group to have better scores vs. control at followup: weighted mean difference of 7.73 (95% CI 3.58 to 11.87). In 5 trials counseling improved overall functioning vs. controls on various measures (49 to 80% improved in counseling groups vs. 17 to 58% in controls), while 2 trials reported mixed results with different measures in the same study, 1 trial reported improvement in the control group compared with counseling, and the other 4 trials reported no differences between groups.	Low
Decreased fatigue	12 RCT (n=1,635)	Results were primarily positive, but mixed. In 9 trials counseling significantly decreased fatigue vs. controls on various measures (63 to 76% improved in counseling groups vs. 15-65% in controls), while the other 3 trials reported no differences between groups.	Moderate
Improved quality of life	5 RCT (n=539)	Results were mixed. In 2 trials counseling showed an improvement in quality of life vs. controls on various measures (mean QOLS at 12 weeks: 2.81 vs. 3.26; p=0.02 and mean change in QLI scores from baseline at 12 months: 2.6 vs. 0.6; p<0.05), 1 trial reported better quality of life vs. support but not no treatment, and the other 2 trials reported no differences between groups.	Low
Increased proportion working full- or part-time	2 RCT (n=145)	No significant differences between intervention and control.	Low
Increased hours worked	3 RCT (n=321)	Significantly more hours worked per week for CBT group vs. control (mean 35.57 vs. 24.00; p<0.04) for 1 trial. The other 2 trials reported no significant differences between intervention and no intervention.	Low
Decreased work impairment	2 RCT (n=531)	Significant improvement reported in both studies for CBT group on work and social adjustment scale compared with controls (mean at 6 months: 3.3 vs. 5.4; p<0.001 on scale scored with range 0-8; mean at 1 year: 21.0 vs. 24.5; p=0.0001 on scale scored with range 0-45).	Low
Global improvement	3 RCT (n=727)	All 3 trials report better global improvement for CBT vs. control (41 to -70% improved in CBT vs. 15-32% in controls).	Moderate

Key question Outcome	Study design Number of studies (n)*	Findings and direction of effect	Strength of evidence grade
Acclydine vs. placebo			
Improved overall function, decreased fatigue, and increased physical activity (actometer)	1 RCT (n=57)	No significant differences between intervention and placebo.	Insufficient
Improved quality of life, increased days spent at work/school, proportion working full-or part-time	No studies	No studies.	Insufficient
Acetyl-L-carnitine vs. propionyl-L-carnitine vs. combination			
Decreased fatigue	1 RCT (n=89)	Acetyl-L-carnitine had lower fatigue scores at 24 weeks, but propionyl-L-carnitine and the combination group improved more from baseline (p=0.004 and p<0.001, respectively).	Insufficient
Global improvement	1 RCT (n=89)	Significant improvement in propionyl-L-carnitine (63%, p<0.001) and acetyl-L-carnitine (59%, p<0.001) compared with the combination group (37%, p=0.084).	Insufficient
Improved overall function, quality of life, increased days spent at work/school, proportion working full- or part-time	No studies	No studies.	Insufficient
Pollen extract vs. placebo			
Decreased fatigue	1 RCT (n=22)	Significant improvement on fatigue scores in the pollen group compared with placebo at 3 months (-0.43 vs. -0.18, p<0.05).	Insufficient
Improved quality of life	1 RCT (n=22)	Significant improvement in quality of life scores in the pollen group compared with placebo at 3 months (-1.66 vs. -0.21; p<0.01).	Insufficient
Improved overall function, increased days spent at work/school, proportion working full- or part-time	No studies	No studies.	Insufficient
Low sugar/low yeast diet vs. healthy eating			
Decreased fatigue, improved quality of life	1 RCT (n=39)	No significant differences between interventions.	Insufficient
Improved overall function, increased days spent at work/school, proportion working full- or part-time	No studies	No studies.	Insufficient

Key question Outcome	Study design Number of studies (n)*	Findings and direction of effect	Strength of evidence grade
Distant healing vs. no treatment			
Improved overall function	1 RCT (n=409)	Significant improvement on functioning scores for those who were blinded to the treatment compared with those who were not blinded to the treatment (covariance analysis effect for blinded vs. unblinded treatment: -1.54 [SE 0.70] 95% CI -2.91 to -0.18). No other significant differences between intervention and no treatment.	Insufficient
Decreased fatigue, improved quality of life, increased days spent at work/school, proportion working full- or part-time	No studies	No studies.	Insufficient
Homeopathy vs. placebo			
Decreased fatigue	1 RCT (n=89)	Significantly better scores on MFI-20 for placebo group compared with intervention at 6 months (mean: 2.70 vs. 1.35, p=0.04).	Insufficient
Improved overall function	1 RCT (n=89)	No significant differences between intervention and placebo.	Insufficient
Improved quality of life, increased days spent at work/school, proportion working full-or part-time	No studies	No studies.	Insufficient
Melatonin vs. phototherapy			
Improved overall function and decreased fatigue	1 RCT crossover design (n=30)	No significant differences between interventions.	Insufficient
Improved quality of life, increased days spent at work/school, proportion working full-or part-time	No studies	No studies.	Insufficient
Home orthostatic training vs. sham home orthostatic training			
Improved overall function	1 RCT (n=36)	No significant differences between interventions.	Insufficient
Decreased fatigue, improved quality of life, increased days spent at work/school, proportion working full- or part-time	No studies	No studies.	Insufficient

Key question Outcome	Study design Number of studies (n)*	Findings and direction of effect	Strength of evidence grade
Qigong exercise vs. no qigong exercise			
Improved overall function	1 RCT (n=52)	Significantly better SF-12 physical functioning scores for qigong exercise compared with no exercise at 4 months (mean: 42.7 vs. 35.7, p=0.001).	Insufficient
Decreased fatigue	1 RCT (n=52)	Significantly better Chalder Fatigue Scale scores in exercise group compared with no exercise group at 4 months (mean total: 21.6 vs. 32.1, p<0.001; mean physical fatigue subscale: 12.9 vs. 20.3, p<0.001; mean mental fatigue subscale: 8.8 vs. 11.9, p=0.012).	Insufficient
Improved quality of life, increased days spent at work/school, proportion working full-or part-time	No studies	No studies.	Insufficient
GET vs. no treatment or flexibility/relaxation therapy or adaptive pacing			
Improved overall function	4 RCT (n=619)	Results from 3 studies that used the SF-36 physical functioning subscale were pooled, there was a significant effect for the intervention group to have better scores vs. control at followup: weighted mean difference 10.29 (95% CI 6.71 to 13.86).	Moderate
Decreased fatigue	4 RCT(n=619)	Significantly better Chalder Fatigue Scale scores reported for exercise groups compared with controls in 3 of the studies:Mean total: 13.91 vs. 24.41; p=0.02, physical fatigue scores: 7.91 vs. 14.27; p=0.02; and mental fatigue scores: 6.00 vs. 10.14; p=0.03 at 12 weeks; mean total: 20.5 vs. 27.4; p=0.004 at 12 weeks; and mean difference in change from baseline from adaptive pacing: -2.5; 95% CI -4.2 to -0.9; p=0.0059 and no treatment: -3.4; 95% CI -5.0 to -1.8; p=0.0001.1 study reported no differences between groups.	Low
Increased proportion working full- or part-time	1 RCT (n=59)	More in the exercise group were working at 1 year compared with control (66% vs. 39%; 95% CI 9% to 44%).	Insufficient
Decreased work impairment	1 RCT (n=475)	Significant improvement reported for exercise group on work and social adjustment scale compared with adaptive pacing and no treatment at 1 year (20.5 vs. 24.5 vs. 23.9; p=0.0004 and p<0.001, respectively).	Low

Key question Outcome	Study design Number of studies (n)*	Findings and direction of effect	Strength of evidence grade
Global improvement	3 RCT (n=583)	Significantly more improvement reported in exercise groups (31% and 54%) compared with controls (7%, p=0.05 and 24%, p=0.04). RR 1.54 (95% CI 1.26 to 1.89)	Moderate
Improved quality of life, increased days spent at work/school	No studies	No studies.	Insufficient
GET ± fluoxetine vs. fluoxetine ± placebo			
Improved overall function	1 RCT (n=136)	Significant improvement for exercise groups (either alone or combination) on functional work capacity at 26 weeks (mean change from baseline: 1.9; 95% CI 0.15 to 3.69; p=0.03) compared with other groups.	Low
Decreased fatigue	1 RCT (n=136)	Significantly more individuals in exercise groups (either alone or combination) did not meet the threshold of "caseness" for fatigue on Chalder Fatigue Scale (18% for both exercise groups and 6% for both other groups; p=0.025).	Low
Increased days spent at work/school and proportion working full- or part-time	No studies	No studies.	Insufficient
Face-to-face CBT vs. telephone CBT			
Clinical global improvement	1 RCT (n=65)	More individuals rated as much better or very much better in face-to-face group compared with telephone group (6 months: 60% vs. 40%; p=NR and 12 months: 57% vs. 55%; p=NR).	Insufficient
Improved overall function, decreased fatigue and work impairment	1 RCT (n=65)	No significant differences between interventions.	Insufficient
Quality of life, days spent at work/school, proportion working full-or part-time	No studies	No studies.	Insufficient
CBT + GET vs. usual care			
Improved overall function, and decreased fatigue	1 RCT (n=115)	No significant differences between intervention and control.	Insufficient
Improved quality of life, decreased work impairment, increased days spent at work/school, proportion working full-or part-time	No studies	No studies.	Insufficient

Key question Outcome	Study design Number of studies (n)*	Findings and direction of effect	Strength of evidence grade
KQ2b. What are the harms of therapeutic interventions for patients with ME/CFS and how do they vary by patient subgroups?			
Galantamine vs. placebo	1 RCT (n=434)	90% (389/434) reported harms; 23% (88/389) withdrew due to harms; 2% (8/389) in galantamine reported serious harms but none attributed to the study drug; no significant differences reported between groups.	Insufficient
Hydrocortisone vs. placebo	1 RCT (n=70)	More harms reported with hydrocortisone vs. placebo (suppression of adrenal glucocorticoid responsiveness: 12 vs. 0; p<0.001; increased appetite: 17 vs. 8; p=0.02; weight gain: 19 vs. 8; p=0.006; difficulty sleeping: 17 vs. 8; p=0.02); no other significant differences between groups.	Insufficient
Hydrocortisone + fludrocortisone vs. placebo	1 RCT (n=80)	1.3% (1/80) withdrew due to acne and weight gain, no serious harms reported; no other harms data reported.	Insufficient
Immunoglobulin G vs. placebo	1 RCT (n=28)	Significantly more with headaches in immunoglobulin G group vs. placebo (93% vs. 60%; p=0.03); 20% total harms overall; 1 in each group withdrew due to harms; 2 in immunoglobulin G and 3 in placebo developed serious harms.	Insufficient
Rintatolimod vs. placebo	2 RCT (n=324)	Flu-like syndrome, chills, vasodilatation, and dyspnea were more frequent in rintatolimod vs. placebo (p<0.05); no other differences between groups.	Insufficient
Valganciclovir vs. placebo	1 RCT (n=30)	No one withdrew due to harms, 1 in each group developed cancer, deemed unrelated; no other harms data reported.	Insufficient
Isoprinosine vs. placebo	1 RCT (n=15)	No one withdrew due to harms; no other harms data reported.	Insufficient
Fluoxetine vs. placebo	1 RCT (n=68)	More total withdrawals in the fluoxetine group compared with placebo.	Insufficient
CBT/counseling vs. no treatment or support or relaxation or adaptive pacing			
Withdrawals due to harms	1 RCT (n=47)	1 trial reported none withdrew due to harms.	Insufficient
Rates of harms	1 RCT (n=257)	1 trial reported no differences between groups for reported harms.	Insufficient
Total harms	1 RCT (n=471)	1 large trial reported fewer total harms in the CBT group (848) vs. adaptive pacing (949) and no treatment (977), but p=NR. The other study did not report harms by group, but deemed all unrelated to the intervention.	Low

Key question Outcome	Study design Number of studies (n)*	Findings and direction of effect	Strength of evidence grade
Serious harms	2 RCT (n=728)	1 large trial (n=471) reported fewer serious harms in the CBT group per 100 person-years (5.0; 95% CI 2.2 to 9.8) vs. adaptive pacing (10.1; 95% CI 5.8 to 16.3), but was similar to no treatment (4.4; 95% CI1.8 to 9.0). The other trial reported that no serious harms were reported.	Low
Acclydine vs. placebo	No studies	No studies.	Insufficient
Acetyl-L-carnitine vs. propionyl-L-carnitine vs. combination	1 RCT (n=89)	No differences reported between groups for withdrawals due to harms; no other harms data reported.	Insufficient
Pollen extract vs. placebo	No studies	No studies.	Insufficient
Low sugar/low yeast diet vs. healthy eating	No studies	No studies.	Insufficient
Distant healing vs. no treatment	No studies	No studies.	Insufficient
Homeopathy vs. placebo	No studies	No studies.	Insufficient
Melatonin vs. phototherapy	No studies	No studies.	Insufficient
Home orthostatic training vs. sham home orthostatic training	No studies	No studies.	Insufficient
Qigong exercise vs. no qigong exercise	1 RCT (n=52)	No harms were reported by either group, no other harms data provided.	Insufficient
GET vs. no treatment or flexibility/relaxation therapy or adaptive pacing			
Withdrawals due to harms	1 RCT (n=49)	1 trial reported 40% (10/25) of GET group refused to repeat the required fitness test due to feeling initial test was harmful and 1 person withdrew due to a calf injury.	Insufficient
Total harms	2 RCT (n=524)	1 trial reported similar harms in the GET group (992) vs. adaptive pacing (949) and no treatment (977), but p=NR. The other trial reported 2% (1/49) experienced a harm.	Low
Serious harms	1 RCT (n=475)	1 large trial reported similar serious harms in GET group per 100 person-years (10.6; 95% CI 6.2 to 17.0) vs. adaptive pacing (10.1; 95% CI 5.8 to 16.3) but fewer in no treatment (4.4; 95% CI 1.8 to 9.0).	Low
GET vs. fluoxetine vs. combination or placebo	1 RCT (n=136)	11 withdrawals due to medication side effects 13% in fluoxetine group vs. 3% in placebo group; no other harms data reported in study.	Insufficient
Face-to-face CBT vs. telephone CBT	No studies	No studies.	Insufficient
CBT + GET vs. usual care	No studies	No studies.	Insufficient

Key question Outcome	Study design Number of studies (n)*	Findings and direction of effect	Strength of evidence grade
KQ2c. What are the characteristics of responders and non-responders to interventions?			
CBT vs. no treatment			
Baseline differences	1 RCT (n=27)	Significant differences between those who responded to CBT and those who did not on baseline measures of functional impairment on SIP-8 (mean: 1,330 vs. 1,985; p=0.031), daily observed fatigue (mean on scale 0-16: 7.4 vs. 9.7; p=0.023), and daily observed pain (mean on scale 0-16: 4.5 vs. 7.8; p=0.026); but not for hours worked per week (mean: 10.9 vs. 2.6; p=0.062).	Insufficient

* Sample size includes only those analyzed

CBT= Cognitive Behavioral Therapy; CFS = chronic fatigue syndrome; CI= Confidence Interval; CIS= Checklist of Individual Strength; FSS= Fatigue Severity Scale; GET= graded exercise therapy; KPS = Karnofsky Performance Score; KQ= key question; ME = Myalgic encephalomyelitis; MFI-20= Multidimensional Fatigue Inventory; n= sample size; NR= not reported; QLI= Quality of Life Index; RCT = randomized controlled trial; SE= standard error; SF-12= Short Form 12-item Health Survey; SF-36= 36-item Short Form Survey; SIP-8 = Sickness Impact Profile 8-items; SOFA-CFS= Schedule of Fatigue and Angina for CFS scale; vs.= versus.

DISCUSSION

Key Findings and Strength of Evidence

Twenty-eight studies contributed to our understanding of diagnostic methods, diagnostic accuracy or concordance, and harms associated with diagnosis of ME/CFS. Multiple case definitions have been used to define ME/CFS and those that require the symptoms of PEM and neurological and autonomic manifestations appear to represent a smaller but more involved subset of the broader population. The artificial neural network test was found to have good sensitivity, specificity, and accuracy for diagnosing ME/CFS (95%, 85% and 90% respectively) based on one small derivation and one small validation cohort. The SOFA-CFS, and certain SF-36 subscales or combination of subscales shows moderate ability to discriminate between patients with ME/CFS compared with healthy controls. None however have been adequately tested in a large population of a broad spectrum of patients with diagnostic uncertainty to determine validity and generalizability. Other tests including serum parameters and cardiopulmonary function and recovery, have been insufficiently tested in broad populations to determine utility. We did not find evidence on how diagnostic tests for ME\CFS vary by subgroups of the population or studies on which related conditions should be ruled out prior to making a ME\CFS diagnosis. Evidence suggests that carrying an ME/CFS diagnosis is associated with perceived stigma, financial instability, difficulty in social interactions and relationships, and a greater risk of receiving a psychiatric diagnosis.

The challenge of diagnosing ME/CFS is evident by the diverse approaches of this literature. The lack of a clear etiology for ME/CFS the multisystem involvement of the syndrome, and its overlap with other chronic conditions contribute to the difficulty in diagnosing ME/CFS. Much research in this field focuses on discovering etiologies rather than testing diagnostic strategies. Articles that attempted to define an etiology on the basis of a biochemical marker or a particular physiologic test were not included in this review because the intent of these was to identify an etiology rather than understand how the specific test could distinguish patients that would respond to treatment. In addition to biomarker studies (cell function, immunologic, virologic/bacteriologic, hormonal, etc.), studies identified subgroups on the basis of exercise testing,[39,40] cerebral blood flow as measured by arterial spin labeling,[41] gait kinetics,[42] impaired blood pressure variability/hemodynamic instability,[43,44] bioenergetics (capacity to recover from acidosis),[45] and many others. These studies did not, in general, report diagnostic testing outcomes such as ROC/AUC, sensitivity, specificity, or concordance and were therefore not included in the diagnostic testing section of this review. The studies on serum parameters and cardiopulmonary function/recovery that did meet the inclusion criteria were not adequately tested in a broad spectrum of patients to determine utility for distinguishing patients with ME/CFS compared with other patients with chronic and disabling conditions.

One of the primary limitations in the literature about diagnostic tests was that very few studies included a validation cohort. Instead, these studies primarily evaluated a diagnostic test in a single initial population (a derivation cohort). Derivation studies are a necessary first step when attempting to achieve a valid diagnostic test, but also have inherent methodological problems. They often involve the use of cases and controls – two very distinct populations – in order to determine whether the test can distinguish between those two groups. If the test is capable of distinguishing between two distinct groups, then further testing uses populations that are more closely related (i.e. they have more overlap in terms of symptoms), in order to more rigorously test the diagnostic capability of a particular test. As such, the more rigorous diagnostic testing

studies will include a population for whom the clinician is likely to face diagnostic uncertainty, and then test how well the test does in classifying that population accurately. The studies identified for evaluation of diagnostic tests for ME/CFS fell into three main categories – those that evaluated a diagnostic test or a scale against a reference standard. In this case, the reference standard was typically one or more of several case definitions that have been published (CDC Holmes, 1988 and CDC Fukuda, 1994, Canadian ME/CFS definition, International Consensus Criteria for ME, etc.). A second group of studies evaluated how those case definitions compare with each other, and whether they identify the same or different populations. While this was not a distinct key question, it was felt to shed light on the evolving definition of ME/CFS and the difficulty with identifying a universally acceptable reference standard. The third groups of studies presented here are those that address harms of diagnosis.

ME/CFS is a condition that lacks a universally accepted diagnostic (gold) standard, a criterion that defines the condition. The lack of gold standard poses significant challenges for evaluation of diagnostic tests, and yet this is a situation that arises commonly with conditions that are syndromes. A syndrome is a "combination of symptoms and signs which have been observed to occur together so frequently and to be so distinctive that they constitute a recognizable clinical picture.[46] That is to say that they combination of findings is unusual so as not to be thought of as coincidence. In such situations, the traditional evaluation of a diagnostic test is more challenging. The ME/CFS literature is beginning to test diagnostic strategies but as yet has not presented data that would sufficiently differentiate the diagnosis of ME/CFS from other similar conditions in a population of patients with substantial diagnostic uncertainty. For example, a proposed test might sufficiently distinguish a patient with CFS from one without, but may not be able to distinguish between a patient with CFS and one with depression or rheumatoid arthritis – conditions that a clinician might be considering simultaneously and attempting to rule out in a patient who presents with fatigue. There were no studies that quantitatively compared the diagnostic concordance of two case definitions. Several studies attempted to demonstrate that ME, ME/CFS, and CFS case definitions identify different groups of people. Studies did this by identifying people who met one criteria but not the other.[4,5,47-49] Using this approach, it appears that the case definitions labeled as ME and ME/CFS select a population with more impairment, lower functioning, and higher symptom reporting compared with the case definitions labeled as CFS alone. Other studies compared subjects who met a definition of CFS with subjects who had other disease states and/or those comprising a normal control population.[36-38] As expected, these studies demonstrated CFS subjects have lower functioning and higher symptom burdens than people from the general population.

One systematic review compared case definitions for ME/CFS using a slightly different approach. This review summarized how the prevalence of ME/CFS in a population and the symptom burden for patients varies when using different case definitions.[50] This study attempted to bring some consistency to case definitions for ME/CFS in the absence of a reference standard. Their inclusion criteria were broader than those for this review and similarly, they found that the validation studies were weak and heterogeneous. This group called for the community of ME/CFS researchers to prioritize research on treatments using existing case definitions (of which, they felt the CDC Fukuda, 1994 criteria had the most studies on validation and comparison with other measures, and was thought to be the most appropriate for clinical practice), rather than development of additional new case definitions.[50]

Patients with ME/CFS report feeling stigmatized by their diagnosis in terms of financial stability, work opportunities, perceived judgments on their character, social isolation, and

interactions with the health care system. Compounding this is the substantial burden of misdiagnosis among this patient population. Two studies objectively identified prejudice and stereotypes towards patients with ME/CFS from members of the medical community; medical trainees and mental health practitioners make judgments about a patient's condition based on the name it carries (ME, CFS, or other) and what treatment is being given. While these studies were descriptive and based on survey data, the results suggest valid concerns about the harm of labeling patients with a diagnosis of ME/CFS. These harms may reflect the chronic and disabling nature of this disease, combined with a lack of understanding about the diagnosis among the medical community and uncertainty about the etiology of ME/CFS.

Thirty-six trials contributed to our understanding of the efficacy of interventions to treat ME/CFS. Although eight of the nine pharmacological trials were targeting an underlying pathophysiological dysfunction, most of the other interventions were targeting associated symptoms of the disease. Compared with usual care, support, relaxation, or adaptive pacing, strength of evidence was moderate with CBT for outcomes of fatigue and global improvement, and for GET for measures of function and global improvement. Strength of evidence was low for CBT for outcomes of function, quality of life, and employment, for GET on outcomes of fatigue. GET was also superior to fluoxetine on measures of function and fatigue (low strength), and rintalimod was superior to placebo on measures of exercise performance (low strength). For all other interventions and outcomes, strength of evidence was insufficient because these outcomes were either not reported, the study quality was poor, and/or the sample size was inadequate to provide a useful estimate. Immune modulators suggested potential improvement in symptoms including improved measures of exercise, fatigue, performance, activities of daily living, and reduced use of other medications for relief of ME/CFS symptoms, while one trial of an antiviral showed improved measures of fatigue. The benefit of homeopathy and L-carnitine preparations remain uncertain; improvement was found in some measures but not in other measures of the same outcome. Although adverse effects were not well reported across trials, GET was associated with a high number of reported adverse events and withdrawal rate in several trials.

Determining the efficacy of interventions to treat ME/CFS was limited because most interventions were only evaluated in single studies with significant methodological limitations including small sample size. Additionally, outcomes were assessed using different methods and scales. None of the medications evaluated are FDA approved for this indication and two are not available in the US (rintatolimod, Isoprinosine). Although trials of immune modulating and antiviral medications suggested potential improvement in symptoms, some of these comparisons were of borderline statistical significance and other measures did not differ between groups. These trials used selective inclusion criteria beyond ME/CFS diagnostic criteria that may have enhanced their ability to detect treatment effects. Two trials of rintatolimod included only severely debilitated patients and a trial of valganciclovir included patients with elevated antibody titers. Trials of Isoprinosine, galantamine, hydrocortisone, immunoglobulin G, and fluoxetine indicated no significant improvement compared with placebo.

Trials were generally designed as pilot studies at single centers and were small, with some enrolling fewer than 20 subjects in an arm. Some trials were primarily intended to measure intermediate outcomes, such as natural killer cell-mediated cytotoxicity,[51] and were underpowered for the health outcomes relevant to this systematic review. While several fatigue and function outcomes were based on validated scales and measures, others were not, and the clinical significance of changes in scores over time are not clear.

The rationale for treating patients with medications that have antiviral or immunomodulatory properties is based on the association of ME/CFS with viruses and immunological abnormalities that may underlie or promote its pathogenesis.[13,52-54] Although small trials of acyclovir,[55] immunoglobulin G,[56,57] and Isoprinosine[51] indicated no statistically significant differences between treatment and placebo groups for measures of fatigue, quality of life, or function, two trials of intravenous rintatolimod[58,59] and a trial of oral valganciclovir[60] suggested improvement. These trials differed from the earlier trials by using newer medications and applying selective inclusion criteria for participants that targeted patient subgroups based on clinical history of a likely viral onset of ME/CFS and high antibody titers[60] or severe disability.[58,59] In addition, the two rintatolimod trials were much larger than the others providing stronger statistical power to detect differences between groups. However, these studies are not definitive and are limited by inconsistencies in methods and findings, small numbers, methodological shortcomings, and lack of long term followup.

Consistent with other systematic reviews, both CBT and GET were found to improve symptoms, primarily based on fatigue, function and global improvement outcomes, whereas evidence on other non-pharmacological interventions was inconclusive.[61-64] Results need to be interpreted with caution given that studies often used multiple methods of evaluating their outcomes and several had mixed results on the same measure when comparing different tools. Although some of the studies attempted to measure adherence, inherent inaccuracies exist with self-reporting. Additionally, none of the studies reported time spent supine or other measures of inactivity given that lack of activity has been theorized as one of the factors perpetuating the symptoms.

Harms were not well reported throughout all of the non-pharmacological and CAM interventions. Interestingly, there was an association between degree of function improvement and knowing if one was receiving distant healing therapy or not. Although a previous systematic review suggested that the placebo response in the treatment of CFS is lower (20%) than would be expected (30-50%),[65] the results of the distant healing trial suggest that expectation theory, a patient's expectation and belief of a positive or negative result, may influence the outcome.[66] The harms associated with exercise were generally more implied than specifically stated in the exercise trials.[67-70] In the combination trials, the greatest number of harms were in the GET arm of one trial,[69] lowest adherence was in the exercise arm in another trial,[68] and several trials had greatest withdrawal due to adverse events in the exercise arms.[67,70]

One of the weaknesses of all of the intervention trials is the lack of subgroup analysis based on factors such as clinical features at baseline (extent of PEM, autonomic dysfunction, neurocognitive impairment, etc.), severity of disease, duration of disease, and patient demographics. Effectiveness and/or harms may differ between patient subgroups and given the small sample size of most of the trials, combining all patients may have lessened the effect size. A recent systematic review that compared different case definitions suggest that patients should be classified according to their severity and symptom patterns in order to optimally guide therapy and predict prognosis.[50]

Concerning is that harms of interventions were generally non-specific and poorly described. The higher rates of refusing to repeat physiological testing implies significant harm in at least some of the participants.[67] Several previous studies have found worsening effects with exercise and a survey sponsored by the ME Association found that patients believed that GET made more people worse compared with other treatments.[71,72] One study comparing CBT with cognitive therapy, anaerobic exercise, or relaxation found that those patients who remained within their

energy envelope (avoided overexertion and under exertion by exerting a comfortable range of energy) had a significant improvement in mean fatigue and functioning scores regardless of treatment arm.[73]

Applicability

The applicability of our findings to real-world clinical settings is supported by several features of the body of literature we reviewed. First, we included all recognized case definitions of ME/CFS in order to allow a broad representation of patients. Studies were conducted primarily in the United States or Western Europe and patients had a female predominance which is consistent with clinical practice. Duration of symptoms, while not consistently reported, was broadly represented across studies. The interventions and comparators represented most of the therapeutic modalities commonly used in clinical practice.

There are however several features of this body of evidence that limit its generalizability to the broader population of patients with ME/CFS, including factors surrounding the diagnosis itself. Given that the condition is a syndrome with a constellation of symptoms and lacking a gold standard for diagnostic comparison, it is at inherent risk of bias by the opinion of experts. For example, experts have identified critical features of the condition including PEM, however current methods of testing, comparing, and monitoring this symptom are lacking. Many of the diagnostic studies have also been conducted in a referral based environment and lacking a spectrum of patients, some with and some without the disease. Patients tended to be white middle-aged women and it is unknown if the results are generalizable to other demographic populations. Most of the intervention trials are of small sample size, few treatments being evaluated in greater than one study, and rarely reporting baseline function or severity of illness. Patients from specialty clinics may represent a more severe form of the condition. Patients from rural centers or lacking insurance or finances may not have access to specialty clinics or clinical trials. Additionally, although most trials included patients based on the CDC (Fukuda, 1994) case definition, some included other diagnostic criteria. We elected to include trials using any pre-defined case definition but recognize that some of the earlier criteria, in particular the Oxford (Sharpe, 1991) criteria, could include patients with 6 months of unexplained fatigue and no other features of ME/CFS. This has the potential of inappropriately including patients that would not otherwise be diagnosed with ME/CFS and may provide misleading results. Finally, given the heterogeneity in the outcomes evaluated and the methods of measuring the outcomes of interest, quantitative meta-analysis could rarely be performed, and comparison between studies is limited.

Implications for Clinical and Policy Decisionmaking

The limitations in applicability as well as the limitations of the evidence base make it difficult to draw firm conclusions with implications for clinical practice. Studies surrounding aspects of diagnosis suggest that different case definitions introduce variability in the characteristics of the population identified as having ME/CFS and that some of the case definitions will be more inclusive (including patients with overlapping conditions) whereas others may be more specific but identify a smaller population with more severe forms of the condition. No one tool has been adequately studied or identified to clearly discriminate between patients particularly when there is diagnostic uncertainty.

Most of the evidence available surrounding treatment is insufficient to draw conclusions. Because of limitations in the evidence base, we did not have high confidence in any of the findings from this review, and only had moderate confidence in the benefit of CBT (fatigue and global improvement) and GET (function and global improvement). In clinical practice, treatment

of ME/CFS often involves multiple concurrent therapies but we found few trials that compared one intervention to another or that compared a combination of concurrent therapies with another. The trial on valganciclovir, an antiviral medication pre-selected patients with an inciting febrile event with lymphadenopathy suggests improvement in fatigue in this population of ME/CFS while the trials on immune modulators, that included patients who were severely disabled, found some improvement in exercise capacity. Both CBT and GET showed improvement in most outcomes but the combination of CBT and GET has not been adequately studied (one trial) to determine if this is more effective than a single intervention. Subgroup analysis in the GET trials would help in identifying if there are specific patients who might have greater benefit or experience greater harm. With CAM interventions, homeopathy may have some benefit, however the variability of specific prescriptions between providers limits that ability to determine what components provided benefit or harm. Similarly, L-carnitine preparations showed benefit on some measures of fatigue and not on others. When multiple measures are used for the same outcome, finding benefit on one scale but not on others draws the clinical significance of the findings into question. Other interventions have not been adequately studied (single small trials, heterogeneity in outcomes measured) to guide clinical and policy decision making.

Across all intervention trials, heterogeneity in the population samples (different case definitions used for inclusion), outcomes evaluated, and tools used to measure these outcomes, limited the ability to synthesize data. Acceptance of a single case definition and development of a core outcome set would aide in better studying the interventions to allow for more meaningful guidance for clinicians, policy makers, and patients.

Limitations of the Evidence Base

What are the limitations of the evidence?

The most important potential limitation of our review is that important studies whose findings might influence clinical and policy decision making may not have been identified. We conducted a comprehensive, broadly inclusive search that produced 5,902 study titles and abstracts. Although we excluded non-English language studies and studies published before 1988, we do not believe that important studies of therapies used in current practice were missed; the general consistency of our findings with other systematic reviews provides some assurance that our review was not biased by our selection criteria. Our review focused on diagnostic methods that provided data on a test's utility in identifying patients with ME/CFS (ROC/AUC, sensitivity, specificity, concordance). Other testing strategies were not reviewed and may provide further insight pertaining to methods of identifying patients with ME/CFS. To evaluate the effectiveness and harms of interventions, we elected to include studies of 12 weeks or longer duration due to the cyclical nature of the condition. Notably, often antiviral or antibiotic medications are traditionally prescribed for a shorter duration and would not have been included in our report. To account for this, we performed a concurrent search for antiviral therapies in the treatment of ME/CFS and only found one additional trial of shorter duration that did not change our results.[55] We considered outcomes of overall improvement, fatigue, function, quality of life, and employment, which we considered clinically significant and conducive to the systematic review methodology. Given the breadth of symptoms in ME/CFS, we a priori elected to not review symptom related outcomes except for fatigue. Some interventions may have revealed benefit for other characteristics of ME/CFS and this review would not have identified these outcomes.

There may have been biased reporting of results in the literature such that only selected studies were published and retrievable, and that published studies may have been affected by conflicts of interest, outcome reporting bias or analysis reporting bias. Reporting bias and

conflicts of interest are concerns with any systematic review. We were not able to conduct quantitative analyses to evaluate the possibility of publication bias for our findings because of the heterogeneity across studies in our review, and in many cases the lack of key information needed to perform quantitative syntheses, generally precluded meaningful comparison of effect sizes. Mitigating against the likelihood of publication bias in our review, however, is the fact that the majority of studies in our review were small (most <100 patients, many <50) and most reported no significant effect of the intervention. Publication bias typically results in selective publication of larger studies and/or those with positive findings, and studies biased by conflict of interests would also be more likely to report positive findings. We also conducted gray literature searches to look for unpublished data and did not find evidence of unreported studies. We did not have access to study protocols, such that our assessment of outcome and analysis reporting bias were limited. However, the limited and vague reporting of harms in many studies may suggest outcome reporting bias for these outcomes.

The main limitation of the evidence base in our review was poor study quality. Most trials did not specify randomization method, did not conceal allocation, and did not mask outcomes assessment. Most studies were small and many were underpowered to detect significant differences. Studies were also highly variable in terms of methods used to measure outcomes limiting our ability to combine or compare results across studies.

Future Research

What are the future research needs for definition, diagnosis, and treatment of ME/CFS?

Future studies evaluating the diagnostic capability of instruments for the identification of ME/CFS should include populations that include a broad range of people with a full complement of conditions that require clinical distinction from ME/CFS, such as fibromyalgia and depression. Thus, the ideal diagnostic test for ME/CFS would adequately distinguish between ME/CFS and these conditions. Additionally, studies should report statistics on how well a particular measure distinguishes a group with ME/CFS from a group that does not meet these criteria – using concordance and the net reclassification index. For physiological and metabolic testing, selection of a broader spectrum of patients as a comparative group is needed rather than healthy controls.

To inform clinical practice and policy, it would be ideal if future intervention studies consistently used an agreed upon single case definition to reduce variability in the patient samples. Trials should use multicomponent treatments, larger sample sizes, with power calculations and more rigorous adherence to methodological standards for clinical trials or observational studies,. Given the cyclical nature of the condition, followup periods greater than one year would be optimal to determine the effectiveness over time. Given the plethora of outcome measures development of a set of core outcomes including more patient-centered outcomes such as quality of life, employment, and time spent supine versus active, would help guide research and facilitate future data syntheses. Reporting of information about cointerventions, the timing of studied interventions in relation to other interventions, and adherence to interventions would improve the applicability of study findings. Similarly, stratification of findings by patient characteristics (e.g., baseline severity, comorbidities, demographics) would help determine the applicability of different interventions for specific patients and situations. It is particularly important for future studies to report findings according to the cardinal features of ME/CFS such as PEM, neurocognitive status, and autonomic function as treatment choices may differ for subsets of the population. Clearly reporting harms particularly surrounding exercise testing and treatment for specific subgroups may help identify patients more negatively affected by these interventions.

Conclusions

Multiple case definitions for ME/CFS exist with those that require symptoms of PEM, neurological impairment, and autonomic dysfunction representing a more severe form of the condition. No current diagnostic tool or method has been adequately tested to identify patients when diagnostic uncertainty exists. Although CBT and GET have shown benefit in some measures of fatigue, function, and global improvement, most other interventions have insufficient evidence to direct clinical practice. GET appears to be associated with harms in some patients whereas the negative effects of being given a diagnosis of ME/CFS appear to be more universal.

References

1. Carruthers BM, Jain AK, de Meirleir KL, et al. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: Clinical Working Case Definition, Diagnostic and Treatment Protocols. J Chronic Fatigue Syndr. 2003;11(1): 7-115.

2. Carruthers BM, van de Sande MI, De Meirleir KL, et al. Myalgic encephalomyelitis: International Consensus Criteria. J Intern Med. 2011;270(4): 327-38. PMID: 21777306.

3. Fukuda K, Straus SE, Hickie I, et al. The chronic fatigue syndrome: a comprehensive approach to its definition and study. International Chronic Fatigue Syndrome Study Group. Ann Intern Med. 1994;121(12): 953-9. PMID: 7978722.

4. Jason LA, Brown A, Evans M, et al. Contrasting chronic Fatigue syndrome versus myalgic encephalomyelitis/chronic fatigue syndrome. Fatigue. 2013;1(3)PMID: 23914329.

5. Jason LA, Brown A, Clyne E, et al. Contrasting case definitions for chronic fatigue syndrome, Myalgic Encephalomyelitis/chronic fatigue syndrome and myalgic encephalomyelitis. Eval Health Prof. 2012;35(3): 280-304. PMID: 22158691.

6. Holmes GP, Kaplan JE, Gantz NM, et al. Chronic fatigue syndrome: a working case definition. Ann Intern Med. 1988;108(3): 387-9. PMID: 2829679.

7. Ramsay M. Myalgic Encephalomyelitis and Postviral Fatigue States: The saga of Royal Free disease. 1st ed. London: Gower Medical Publishing; 1986.

8. Jason LA, Taylor RR, Plioplys S, et al. Evaluating attributions for an illness based upon the name: chronic fatigue syndrome, myalgic encephalopathy and Florence Nightingale disease. Am J Community Psychol. 2002;30(1): 133-48. PMID: 11928774.

9. Komaroff AL, Cho TA. Role of infection and neurologic dysfunction in chronic fatigue syndrome. Semin Neurol. 2011;31(3): 325-37. PMID: 21964849.

10. Morris G, Maes M. A neuro-immune model of Myalgic Encephalomyelitis/Chronic fatigue syndrome. Metab Brain Dis. 2013;28(4): 523-40. PMID: 22718491.

11. Hempel S, Chambers D, Bagnall AM, et al. Risk factors for chronic fatigue syndrome/myalgic encephalomyelitis: a systematic scoping review of multiple predictor studies. Psychol Med. 2008;38(7): 915-26. PMID: 17892624.

12. Chew-Graham C, Dowrick C, Wearden A, et al. Making the diagnosis of Chronic Fatigue Syndrome/Myalgic Encephalitis in primary care: a qualitative study. BMC Fam Pract. 2010;11: 16. PMID: 20178588.

13. Afari N, Buchwald D. Chronic fatigue syndrome: a review. Am J Psychiatry. 2003;160(2): 221-36. PMID: 12562565.

14. Anderson G, Maes M, Berk M. Biological underpinnings of the commonalities in depression, somatization, and Chronic Fatigue Syndrome. Med Hypotheses. 2012;78(6): 752-6. PMID: 22445460.

15. Reyes M, Nisenbaum R, Hoaglin DC, et al. Prevalence and incidence of chronic fatigue syndrome in Wichita, Kansas. Arch Intern Med. 2003;163(13): 1530-6. PMID: 12860574.

16. Reeves WC, Jones JF, Maloney E, et al. Prevalence of chronic fatigue syndrome in metropolitan, urban, and rural Georgia. Popul Health Metr. 2007;5: 5. PMID: 17559660.

17. Johnston S, Brenu EW, Staines D, et al. The prevalence of chronic fatigue syndrome/ myalgic encephalomyelitis: a meta-analysis. Clin Epidemiol. 2013;5: 105-10. PMID: 23576883.

18. Crawley EM, Emond AM, Sterne JA. Unidentified Chronic Fatigue Syndrome/myalgic encephalomyelitis (CFS/ME) is a major cause of school absence: surveillance outcomes from school-based clinics. BMJ Open. 2011;1(2): e000252. PMID: 22155938.

19. Rangel L, Garralda ME, Levin M, et al. The course of severe chronic fatigue syndrome in childhood. J R Soc Med. 2000;93(3): 129-34. PMID: 10741312.

20. Jason LA, Porter N, Hunnell J, et al. A natural history study of chronic fatigue syndrome. Rehabil Psychol. 2011;56(1): 32-42. PMID: 21401284.

21. Cairns R, Hotopf M. A systematic review describing the prognosis of chronic fatigue syndrome. Occup Med. 2005;55(1): 20-31. PMID: 15699087.

22. Center for Drug Evaluation and Research. The Voice of the Patient: A series of reports from the U.S. Food and Drug Administration's (FDA's) Patient-Focused Drug Development Initiative. Chronic Fatigue Syndrome and Myalgic Encephalomyelitis: U.S. Food and Drug Administration. 2013. Availible at: http://www.fda.gov/downloads/ForIndustry/UserFees/PrescriptionDrugUserFee/UCM368806.pdf. Accessed March 22, 2014.

23. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Publication No. 10(13)-EHC063-EF. Rockville (MD) :Agency for Healthcare Research and Quality. January 2014. Availible at: www.effectivehealthcare.ahrq.gov.

24. Smith MEB, Nelson N, Haney E, et al. Diagnosis and treatment of myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS).

PROSPERO 2014:CRD42014009779 2014. Availible at: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42014009779. Accessed July 3, 2014.

25. Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. BMJ. 1994;309(6948): 188. PMID: 8044101.

26. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem. 1993;39(4): 561-77. PMID: 8472349.

27. Hosmer Jr DW, Lemeshow S. Applied logistic regression: John Wiley & Sons; 2004.

28. Methods Guide for Medical Test Reviews. AHRQ Publication No. 12-EC017. Rockville, MD: Agency for Healthcare Research and Quality; June 2012. www.effectivehealthcare.ahrq.gov/reports/final.cfm. Also published as a special supplement to the Journal of General Internal Medicine, July 2012. PMID: 22834019.

29. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. Ann Intern Med. 2011;155(8): 529-36. PMID: 22007046.

30. Harris RP, Helfand M, Woolf SH, et al. Current methods of the US Preventive Services Task Force: a review of the process. Am J Prev Med. 2001;20(3 Suppl): 21-35. PMID: 11306229.

31. Atkins D, Chang SM, Gartlehner G, et al. Assessing applicability when comparing medical interventions: AHRQ and the Effective Health Care Program. J Clin Epidemiol. 2011;64(11): 1198-207. PMID: 21463926.

32. Fu R, Gartlehner G, Grant M, et al. Conducting quantitative synthesis when comparing medical interventions: AHRQ and the Effective Health Care Program. J Clin Epidemiol. 2011;64(11): 1187-97. PMID: 21477993.

33. Sutton A, Abrams K, Jones DR, et al. Methods for Meta-Analysis in Medical Research. Chichester, UK: John Wiley &Sons, Inc; 2000.

34. Sutton GC. 'Too tired to go to the support group': a health needs assessment of myalgic encephalomyelitis. J Public Health Med. 1996;18(3): 343-9. PMID: 8887847.

35. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21(11): 1539-58. PMID: 12111919.

36. Katon WJ, Buchwald DS, Simon GE, et al. Psychiatric illness in patients with chronic fatigue and those with rheumatoid arthritis. J Gen Intern Med. 1991;6(4): 277-85. PMID: 1890495.

37. Komaroff AL, Fagioli LR, Doolittle TH, et al. Health status in patients with chronic fatigue syndrome and in general population and disease comparison groups. Am J Med. 1996;101(3): 281-90. PMID: 8873490.

38. Aslakson E, Vollmer-Conna U, White PD. The validity of an empirical delineation of heterogeneity in chronic unexplained fatigue. Pharmacogenomics. 2006;7(3): 365-73. PMID: 16610947.

39. Vanness JM, Snell CR, Strayer DR, et al. Subclassifying chronic fatigue syndrome through exercise testing. Med Sci Sports Exerc. 2003;35(6): 908-13. PMID: 12783037.

40. Sisto SA, LaManca J, Cordero DL, et al. Metabolic and cardiovascular effects of a progressive exercise test in patients with chronic fatigue syndrome. Am J Med. 1996;100(6): 634-40. PMID: 8678084.

41. Biswal B, Kunwar P, Natelson BH. Cerebral blood flow is reduced in chronic fatigue syndrome as assessed by arterial spin labeling. J Neurol Sci. 2011;301(1-2): 9-11. PMID: 21167506.

42. Boda WL, Natelson BH, Sisto SA, et al. Gait abnormalities in chronic fatigue syndrome. J Neurol Sci. 1995;131(2): 156-61. PMID: 7595641.

43. Frith J, Zalewski P, Klawe JJ, et al. Impaired blood pressure variability in chronic fatigue syndrome--a potential biomarker. QJM. 2012;105(9): 831-8. PMID: 22670061.

44. Naschitz JE, Sabo E, Naschitz S, et al. Hemodynamics instability score in chronic fatigue syndrome and in non-chronic fatigue syndrome.[Erratum appears in Semin Arthritis Rheum. 2003 Apr;32(5):343 Note: Madelain, Fields [corrected to Fields, Madeline]; Hillel, Isseroff [corrected to Isseroff, Hillel]]. Semin Arthritis Rheum. 2002;32(3): 141-8. PMID: 12528078.

45. Jones DEJ, Hollingsworth KG, Jakovljevic DG, et al. Loss of capacity to recover from acidosis on repeat exercise in chronic fatigue syndrome: a case-control study. Eur J Clin Invest. 2012;42(2): 186-94. PMID: 21749371.

46. Scadding JG. Diagnosis: the clinician and the computer. Lancet. 1967;2(7521): 877-82. PMID: 4168324.

47. Jason LA, Torres-Harding SR, Taylor RR, et al. A comparison of the 1988 and 1994 diagnostic criteria for chronic fatigue syndrome. J Clin Psychol Med Settings. 2001;8(4): 337-43.

48. Brown AA, Jason LA, Evans MA, et al. Contrasting case definitions: The ME International Consensus Criteria vs. the Fukuda et al. CFS criteria. N Am J Psychol. 2013;15(1): 103-20.

49. Van Hoof E, De Meirleir K. Chronic Fatigue Syndrome and Myalgic Encephalomyletis: Are Both Conditions on the Same Continuum? N Am J Psychol. 2005;7(2): 189-204.

50. Brurberg KG, Fonhus MS, Larun L, et al. Case definitions for chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME): a systematic review. BMJ Open. 2014;4(2): e003973. PMID: 24508851.

51. Diaz-Mitoma F, Turgonyi E, Kumar A, et al. Clinical improvement in chronic fatigue syndrome is associated with enhanced natural killer cell-mediated cytotoxicity: The results of a pilot study with Isoprinosine. J Chronic Fatigue Syndr. 2003;11(2): 71-93.

52. Komaroff AL, Buchwald DS. Chronic fatigue syndrome: an update. Annu Rev Med. 1998;49: 1-13. PMID: 9509246.

53. Komaroff AL. Is human herpesvirus-6 a trigger for chronic fatigue syndrome? J Clin Virol. 2006;37 Suppl 1: S39-46. PMID: 17276367.

54. Devanur LD, Kerr JR. Chronic fatigue syndrome. J Clin Virol. 2006;37(3): 139-50. PMID: 16978917.

55. Straus SE, Dale JK, Tobi M, et al. Acyclovir treatment of the chronic fatigue syndrome. Lack of efficacy in a placebo-controlled trial. N Engl J Med. 1988;319(26): 1692-8. PMID: 2849717.

56. Peterson PK, Shepard J, Macres M, et al. A controlled trial of intravenous immunoglobulin G in chronic fatigue syndrome. Am J Med. 1990;89(5): 554-60. PMID: 2239975.

57. Lloyd A, Hickie I, Wakefield D, et al. A double-blind, placebo-controlled trial of intravenous immunoglobulin therapy in patients with chronic fatigue syndrome. Am J Med. 1990;89(5): 561-8. PMID: 2146875.

58. Strayer DR, Carter WA, Brodsky I, et al. A controlled clinical trial with a specifically configured RNA drug, poly(I) midline dot poly(C12U), in chronic fatigue syndrome. Clin Infect Dis. 1994;18(SUPPL. 1): S88-S95. PMID: 8148460.

59. Strayer DR, Carter WA, Stouch BC, et al. A double-blind, placebo-controlled, randomized, clinical trial of the TLR-3 agonist rintatolimod in severe cases of chronic fatigue syndrome. PLoS ONE. 2012;7(3): e31334. PMID: 22431963.

60. Montoya JG, Kogelnik AM, Bhangoo M, et al. Randomized clinical trial to evaluate the efficacy and safety of valganciclovir in a subset of patients with chronic fatigue syndrome. J Med Virol. 2013;85(12): 2101-9. PMID: 23959519.

61. Chambers D, Bagnall A-M, Hempel S, et al. Interventions for the treatment, management and rehabilitation of patients with chronic fatigue syndrome/myalgic encephalomyelitis: an updated systematic review. J R Soc Med. 2006;99(10): 506-20. PMID: 17021301.

62. Whiting P, Bagnall AM, Sowden AJ, et al. Interventions for the treatment and management of chronic fatigue syndrome: a systematic review. JAMA. 2001;286(11): 1360-8. PMID: 11560542.

63. Alraek T, Lee MS, Choi T-Y, et al. Complementary and alternative medicine for patients with chronic fatigue syndrome: a systematic review. BMC Altern Med. 2011;11: 87. PMID: 21982120.

64. Kim J-E, Hong K-E, Kim H-J, et al. An open-label study of effects of acupuncture on chronic fatigue syndrome and idiopathic chronic fatigue: study protocol for a randomized controlled trial. Trials. 2013;14: 147. PMID: 23693129.

65. Cho HJ, Hotopf M, Wessely S. The placebo response in the treatment of chronic fatigue syndrome: a systematic review and meta-analysis. Psychosom Med. 2005;67(2): 301-13. PMID: 15784798.

66. Walach H, Bosch H, Lewith G, et al. Effectiveness of distant healing for patients with chronic fatigue syndrome: a randomised controlled partially blinded trial (EUHEALS). Psychother Psychosom. 2008;77(3): 158-66. PMID: 18277062.

67. Moss-Morris R, Sharon C, Tobin R, et al. A randomized controlled graded exercise trial for chronic fatigue syndrome: outcomes and mechanisms of change. J Health Psychol. 2005;10(2): 245-59. PMID: 15723894.

68. Jason LA, Torres-Harding S, Friedberg F, et al. Non-pharmacologic interventions for CFS: a randomized trial. J Clin Psychol Med Settings. 2007;14(4): 275-96.

69. White PD, Goldsmith KA, Johnson AL, et al. Comparison of adaptive pacing therapy, cognitive behaviour therapy, graded exercise therapy, and specialist medical care for chronic fatigue syndrome (PACE): a randomised trial. Lancet. 2011;377(9768): 823-36. PMID: 21334061.

70. Wearden AJ, Morriss RK, Mullis R, et al. Randomised, double-blind, placebo-controlled treatment trial of fluoxetine and graded exercise for chronic fatigue syndrome.[Erratum appears in Br J Psychiatry 1998 Jul;173:89]. Br J Psychiatry. 1998;172: 485-90. PMID: 9828987.

71. Cooper L. Report on Survey of Members of Local ME Groups. Bristol, UK: Action for ME. 2000. Availible at: http://www.actionforme.org.uk/Resources/Action%20for%20ME/Documents/get-informed/Local%20Group%20Survey%202001%20Lesley%20Cooper.PDF. Accessed July 30, 2014.

72. Edmonds M, McGuire H, Price J. Exercise therapy for chronic fatigue syndrome. Cochrane Database Syst Rev. 2004(3): CD003200. PMID: 15266475.

73. Jason L, Benton M, Torres-Harding S, et al. The impact of energy modulation on physical functioning and fatigue severity among patients with ME/CFS. Patient Educ Couns. 2009;77(2): 237-41. PMID: 19356884.

Introduction

Background

Myalgic encephalomyelitis (ME) and/or chronic fatigue syndrome (CFS) is a condition characterized by chronic and disabling fatigue as well as various additional manifestations including pain, sleep disturbance, neurological and cognitive changes, motor impairment, and altered immune and autonomic responses.[1-3] Experts consider post-exertional malaise (PEM) and memory or concentration problems critical components.[4]

The term ME was first used in the 1930s after an outbreak of neuromyesthenia and CFS was first coined in the 1980s.[5-7] Attempts to describe the condition based on possible underlying etiologies led to additional terms including post viral fatigue syndrome and chronic fatigue immune dysfunction syndrome.[1,3,5,6] The most recent international consensus report advocates moving away from the term CFS in favor of ME to better reflect an underlying pathophysiology involving widespread inflammation and neuropathology, and to embrace the two terms as synonymous.[2] However, others believe that ME is a subset of CFS and represents a more severe form of the same disease.[4] Some feel that the lack of specificity surrounding the name, CFS, may delegitimize and negatively characterize the condition, and stigmatize patients.[8] For this review, ME and CFS will be used synonymously (ME/CFS) and will include the populations(s) studied under either of these terms, recognizing that issues regarding terminology are currently unresolved.

Uncertainty persists regarding the etiology and whether the condition reflects a single pathologically discrete syndrome, subsets of the same illness, or a nonspecific condition shared by other disease entities. Some suggest that an inciting event triggers an immune response and promotes immune and/or neuroendocrine dysregulation that perpetuates the body's response and symptom experience that becomes ME/CFS.[9,10] Viral etiologies have been predominantly studied based on the observation that the majority of patients report a sudden onset of symptoms associated with a preceding febrile illness and enlarged lymph nodes. However, no specific virus or other infectious agent has been identified, and not all patients experience a preceding febrile illness.[9] Numerous studies have attempted to identify risk factors for developing ME/CFS but a systematic review in 2008 of 11 studies that assess multiple predictors found no evidence of any definitive factors.[11] This review is not intended to address the question of etiology nor underlying factors that lead to the onset or perpetuation of ME/CFS but rather to focus on the diagnosis and treatment of this syndrome.

Currently, diagnosing a patient with ME/CFS relies on the use of a set of clinical criteria (case definitions) to distinguish ME/CFS from other conditions that may also present with fatigue. There are currently eight published case definitions that have evolved since the first one was published by the Centers for Disease Control and Prevention (CDC) in 1988,[6] all of the definitions include persistent fatigue not attributable to a known underlying medical condition, as well as additional clinical signs and symptoms that do not all need to be present to establish the diagnosis.[4] As yet, none of them has been accepted as a reliable "gold standard". As with other medical syndromes that involve a multitude of symptoms and lack a definitive diagnostic test, differentiating one disease state from another similar or overlapping condition becomes a challenge. Some clinicians are reluctant to diagnose ME/CFS believing that the diagnosis will harm the patient or that the patient be inappropriately labeled.[12] This makes the prevalence of

ME/CFS difficult to assess.[13,14] The CDC reported a U.S. prevalence rate of 0.3 percent corresponding to over 1,000,000 adults.[15] By using different case definitions, the rate may be as high as 2.5 percent.[5,16] A recent systematic review found that when using the same case definition (CDC Fukuda, 1994), the prevalence was higher when determined by self report (3.28%; 95% confidence interval [CI], 2.24 to 4.33) compared with clinical assessment (0.76%; 95% CI, 0.23 to 1.29).[17] ME/CFS is more common among women with the average age of diagnosis between 30 and 40 years.[13] Childhood ME/CFS is uncommon and although symptoms may be similar, prognosis appears to be different.[18,19] Although the natural history is not well studied, symptoms and disability in adults tend to persist over time,[20] and although 40 percent (8-63%) of adult patients improve over time, only 5 percent (0-31%) fully recover[21] in contrast to childhood studies that suggest that over 50 percent of patients will recover within 6 months.[19] Economic impact is considerable with most adult patients never returning to work.[9,21]

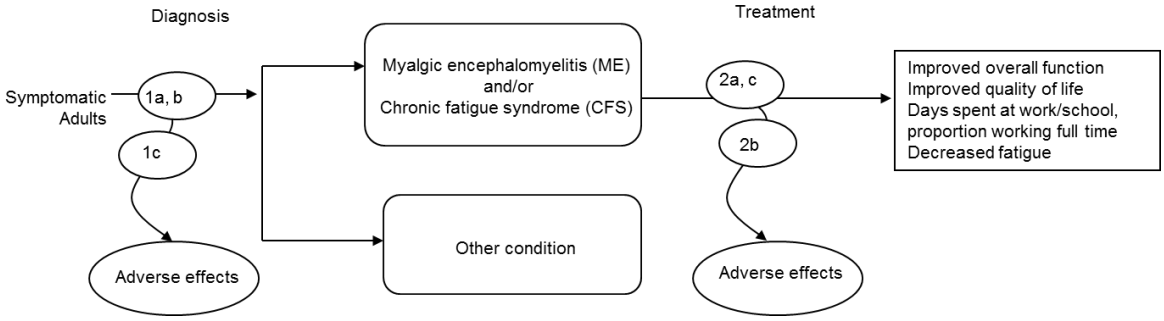
Currently there are no medications for the treatment of ME/CFS approved by the U.S. Food and Drug Administration (FDA), but many have been used 'off-label' (without review and approval), and some have been obtained from other countries and are not currently approved for any indication in the United States (i.e., Isoprinosine, rintatolimod). In a survey by the FDA, patients identified treatments that fell into two broad categories, those intended to treat the underlying cause of the disease and those targeting specific symptoms.[22] Medications to treat underlying causes include immune modulators and antiviral and antibiotic medications. Interventions targeting symptoms include medications to treat pain, fatigue, autonomic dysfunction, and sleep dysfunction, and non-drug therapies such as yoga, exercise techniques, counseling on pacing strategies, and mental exercises.[22] In practice, there are wide variations in the clinical management of patients, and many patients receive a multifaceted approach to treatment.

The variable symptomatology of ME/CFS, lack of a clearly identifiable etiology and disease process, and no accepted diagnostic tests or treatments have challenged researchers and clinicians in their attempts to better understand the condition and treat patients. This systematic review was commissioned by the Office of Disease Prevention (ODP) at the National Institutes of Health (NIH) to inform a Pathways to Prevention Workshop on ME/CFS. This review summarizes the research on diagnosis and treatment of ME/CFS including the methods and criteria used to diagnose ME/CFS, their utility in clinical practice, the harms associated with carrying a diagnosis of MECFS, and the evidence on treatment effectiveness and associated harms. It identifies areas of future research needed to better inform the diagnostic process and treatment strategies. This report is not intended to be used or likely to be useful to develop criteria for disability or insurance.

Scope of Review and Key Questions

This topic was nominated for review by the NIH and focuses on diagnosis and treatment for ME/CFS. The analytic framework (**Figure 1**) and key questions used to guide this review are shown below. The analytic framework shows the target populations, interventions, and health outcomes we examined, with numbers corresponding to the key questions.

Figure 1. Analytic framework



The following key questions are the focus of the report:
Key Question 1. What methods are available to clinicians to diagnose ME/CFS and how do the use of these methods vary by patient subgroups?

Key Question1a. What are widely accepted diagnostic methods and what conditions are required to be ruled out or excluded before assigning a diagnosis of ME/CFS ?

Key Question 1b.What is the accuracy and concordance of diagnostic methods?

Key Question 1c. What harms are associated with diagnosing ME/CFS?

Key Question 2a. What are the (a) benefits and (b) harms of therapeutic interventions for patients with ME/CFS and how do they vary by patient subgroups?

Key Question 2c. What are the characteristics of responders and non-responders to interventions?

Methods

This systematic review follows the methods of the Agency for Healthcare Research and Quality (AHRQ) Methods Guide for Effectiveness and Comparative Effectiveness Reviews (hereafter "AHRQ Methods Guide").[23]

Topic Development and Refinement

The initial key questions were developed in conjunction with the NIH and AHRQ, and revised with input from a Technical Expert Panel (TEP) convened for this report. The TEP consisted of experts in six disciplines and two patients who disclosed no conflicts of interest that precluded participation. The TEP did not contribute to reviewing the evidence or writing of the report.

With input from the TEP, the NIH, and AHRQ, the final protocol was developed and posted on the AHRQ web site on May 1, 2014 at: http://effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?productid=1906&pageaction=displayproduct. The protocol was also registered in the PROSPERO international database of prospectively registered systematic reviews.[24]

Literature Search Strategy

A research librarian conducted searches in Ovid MEDLINE (1988 to November 2013), PsycINFO (1988 to January 2014), the Cochrane Central Register of Controlled Trials and Cochrane Database of Systematic Reviews (through November 2013), and the Database of Abstracts of Reviews of Effects and the National Health Sciences Economic Evaluation Database (through the 4th quarter 2013) see **Appendix A** for full search strategies. Searches were supplemented with hand-searches of reference lists of relevant studies. In addition, scientific information packets were requested from drug and device manufacturer who potentially had data on the use of medications or devices for ME or CFS, who had the opportunity to submit data using the portal for submitting scientific information packets on the Effective Health Care Program Web site. Seventeen submissions were received. Library searches will be updated while the draft report is under external review.

Process for Study Selection

Criteria for inclusion and exclusion of studies was developed based on the key questions and the populations, interventions, comparators, outcomes, timing, types of studies, and setting (PICOTS) approach (**Appendix B**). Papers were selected for review if they were about diagnosis or treatment of ME or CFS in adult populations, were relevant to a key question, and met the prespecified inclusion criteria. Studies of nonhuman subjects and studies with no original data were excluded. Abstracts were reviewed by two investigators for inclusion for each key question. Full-text articles were obtained for all studies that either investigator identified as potentially meeting inclusion criteria. Two investigators independently reviewed all full-text articles for final inclusion. Inclusion was restricted to English language articles. A list of the included studies can be found in **Appendix C**; excluded studies can be found in **Appendix D**, with primary reasons for exclusion. Discrepancies were resolved through discussion and consensus, with a third investigator making the final decision if necessary.

Population and Conditions of Interest

For Key Question 1, symptomatic adults, ages 18 years or older, with fatigue were included. For Key Question 2, adults, ages 18 years or older, diagnosed with ME, CFS, or both by fulfilling criteria from at least one of the case definitions and without another underlying diagnosis were included. To minimize heterogeneity in patient populations, we did not include studies in which patients who may have met criteria for ME/CFS were included as part of a broader grouping of an overlapping condition (i.e., depression, fibromyalgia).

Interventions and Comparisons/Study Designs

For Key Question 1, studies were included if they evaluated any diagnostic test and compared it to a reference standard. Because there is no single accepted definition for ME/CFS and therefore no 'gold standard', any of the eight case definitions published since 1988 were used. Case definitions were reviewed in order to interpret studies that defined populations according to different definitions and studies that compared case definitions were also included. Measures of diagnostic accuracy and concordance were considered. Diagnostic accuracy is a measure of how well the model can separate those who do and do not have the disease of interest and is measured by the model's concordance statistic or c-stat. The c-stat is determined by the area under the receiver operator curve, a plot of sensitivity (true positive rate) versus 1-specificity (false-

positive rate). Perfect discrimination is a c-stat of 1.0 and occurs when all cases attain higher risk scores than all non-cases. A c-stat of 0.5 would result from chance alone. An acceptable level of discrimination is considered as ≥0.70 and <0.80, excellent ≥0.80 and <0.90, and outstanding ≥0.90.[25] Concordance refers to how well two tests agree. For harms of diagnosis, studies that evaluated harms by surveys, qualitative interviews, or trials designed to identify perceptions of diagnosis or treatment for ME/CFS.

For Key Question 2, randomized trials comparing medication management (immune modulators, beta blockers, antidepressants, anxiolytics, stimulants, other), counseling and behavior therapy, graded exercise programs, and complementary and alternative medicine (CAM) approaches (acupuncture, relaxation, massage, other) with placebo, no treatment, usual care, or other active interventions, including combination therapies and head-to-head trials were included. For harms, cohort studies with control groups were also included.

Outcomes

For Key Question 1, outcomes of diagnostic accuracy or concordance were considered including sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio, negative likelihood ratio, C-statistic, receiver operator curve (ROC) and area under curve (AUC), net reclassification index, concordance, and any potential harm from diagnosis (such as psychological harms, labeling, risk from diagnostic test, misdiagnosis).

For Key Question 2, outcomes were included if patient-centered and included overall function (i.e., 36-item Short Form Survey [SF-36]), quality of life, days spent at work or school, proportion working full- or part-time, and fatigue (Multidimensional Fatigue Inventory 20-item [MFI-20] or similar). Individual symptom-based outcomes were excluded. Included harms were withdrawals, withdrawals due to harms, and rates of harms due to interventions.

Timing

There was no duration or timing restriction on studies included for Key Question 1.

Intervention studies, for Key Question 2, must have a minimum duration of 12 weeks of treatment to be included given the cyclical nature of the condition characterized by an intermittent pattern of relapse and remission.[26]

Setting

Studies for all Key Questions had to be conducted in a clinical setting or a setting that was generalizable to clinical practice settings. Studies conducted with in-patients or institutionalized individuals were excluded.

Data Extraction and Data Management

The following information was extracted from included studies into evidence tables: study design, setting, inclusion and exclusion criteria, population characteristics (including sex, age, race, and co-morbidities), sample size, duration of followup, attrition, intervention characteristics, case definition used for diagnosis, duration of illness, and results. Data extraction for each study was performed by two investigators: the first investigator extracted the data, and the second investigator independently reviewed the extracted data for accuracy and completeness.

For studies of diagnostic accuracy and concordance, when reported, we extracted relative measures of risk (relative risk [RR], odds ratio [OR], hazards ratio [HR]), ROC, and AUC. The area under the receiver operating characteristic (AUROC), which is based on sensitivities and specificities across a range of test results, is a measure of discrimination, or the ability of a test to distinguish people with a condition from people without the condition.[27,28] An AUROC of 1.0 indicates perfect discrimination, and an AUROC of 0.5 indicates complete lack of discrimination. Interpretation of AUROC values between 0.5 and 1.0 is somewhat arbitrary, but a value of 0.90 to 1.0 has been classified as excellent, 0.80 to <0.90 as good, 0.70 to <0.80 as fair, and <0.70 as poor.

Individual Study Quality Assessment

The quality of each study was assessed based on predefined criteria adapted from methods proposed by the U.S. Preventive Services Task Force. The criteria used are consistent with the approach recommended by AHRQ in the AHRQ Methods Guide.[23] The term "quality" was used rather than the alternate term "risk of bias;" both refer to internal validity. Two investigators independently assessed the quality of each study. Discrepancies were resolved through discussion and consensus, with a third investigator making the final decision if necessary.

To determine the quality of each study evaluating diagnostic tests, we used questions from the AHRQ Methods Guide for Medical Test Reviews and adapted them to improve their clinical relevance to ME/CFS.[29] Quality was based on whether the study evaluated a representative spectrum of patients, whether it enrolled a random or consecutive sample of patients meeting prespecified criteria, whether it used a credible reference standard, whether the same reference standard was applied to all patients, whether the reference standard was interpreted independently from the test under evaluation, and whether thresholds were prespecified.[23,30,31] Descriptive papers that compared diagnostic criteria and reported harms were not quality rated.

The quality of intervention trials was based on the methods used for randomization, allocation concealment, and blinding; the similarity of compared groups at baseline; maintenance of comparable groups; adequate reporting of dropouts, attrition, crossover, adherence, and contamination; loss to followup; the use of intent-to-treat analysis; and ascertainment of outcomes.[23,31]

Following assessment of individual quality criteria, individual studies were rated as "good," "fair," or "poor" quality, as defined below.[23,29]

Good-quality studies are considered likely to be valid. Good-quality studies clearly describe the population, setting, interventions, and comparison groups; use a valid method for allocation of patients to interventions; clearly report dropouts and have low dropout rates; use appropriate methods for preventing bias; assess outcomes blinded to intervention status; and appropriately measure outcomes and fully report results.

Fair-quality studies have some methodological deficiencies, but no flaw or combination of flaws judged likely to cause major bias. The study may be missing information, making it difficult to assess its methods or assess limitations and potential problems. The fair-quality category is broad, and studies with this rating vary in their strengths and weaknesses: the results of some fair-quality studies are likely to be valid, while others are probably *in*valid.

Poor-quality studies have significant flaws that may invalidate the results. They have a serious or "fatal" flaw in design, analysis, or reporting; large amounts of missing information; or discrepancies in reporting. The results of these studies are judged to be at least as likely to reflect flaws in the study design as true effects of the interventions under investigation. Poor-quality

studies were not excluded *a priori*, but they were considered to be the least reliable studies when synthesizing the evidence, particularly when discrepancies between studies were present. For detailed quality assessment criteria see **Appendix E.**

Assessing Research Applicability

Applicability is defined as the extent to which the effects observed in published studies are likely to reflect the expected results when a specific intervention is applied to the population of interest under "real-world" conditions.[23] It is an indicator of the extent to which research included in a review might be useful for informing clinical and/or policy decisions in specific situations. Applicability depends on the particular question and the needs of the user of the review. There is no generally accepted universal rating system for applicability. In addition, applicability depends in part on context. Therefore, a rating of applicability (such as "high" or "low") was not assigned because applicability may differ based on the user of this review. Rather, factors important for understanding the applicability of studies were recorded, such as how similar patients were to the population of interest, how large the sample size was, and the characteristics of the clinical setting.[32] The funding source for treatment trials was also recorded.

Data Synthesis

Results of diagnostic accuracy studies (such as creating summary AUROCs) were not quantitatively pooled due to differences in methods, case definitions, and heterogeneity in the outcomes. Instead, descriptive statistics were used, such as the median sensitivity and specificity at specific cutoffs and reported AUROCs, along with associated ranges, and calculated positive and negative likelihood ratios based on the median sensitivities and specificities. For the results of intervention trials, the appropriateness of meta-analysis was determined by considering the internal validity of the studies and the heterogeneity among studies in design, patient population, interventions, and outcomes. Appropriate measures were chosen based on the type of data for meta-analysis, according to the guidance for the Evidence-based Practice Center program.[33] Random-effects models were used to estimate pooled effects, except when only two studies were available we chose not to pool the results.[34] We calculated pooled RR where the data were reported as proportions of dichotomous outcomes (e.g., proportion with improvement in intervention and control groups). For continuous outcomes, we calculated pooled weighted mean differences using the means and standard deviations (SDs) (e.g., mean change in function based on a scale). The Q statistic and the I-squared statistic (the proportion of variation in study estimates due to heterogeneity) were calculated to assess heterogeneity in effects between studies.[35,36] When statistical heterogeneity was found, we explored the reasons by using subgroup analysis. In meta-analysis, we combined RRs and ORs for such outcomes.

Grading the Body of Evidence for Each Key Question

The overall strength of evidence was assessed for each key question and outcome in accordance with the AHRQ Methods Guide.[23,29] Strength of evidence was based on the overall quality of each body of evidence, the study limitations (graded low, moderate, or high); the consistency of results between studies (graded consistent, inconsistent, or consistency unknown when only one study was available); the directness of the evidence linking the intervention and health outcomes (graded direct or indirect); the precision of the estimate of effect, based on the number and size of studies and CIs for the estimates (graded precise or imprecise); and whether

reporting bias was suspected (graded suspected or undetected). There was no way to formally assess for publication bias due to small number of studies, methodological shortcomings, or differences across studies in designs, measured outcomes, and other factors. For a more detailed description of the categories used see **Appendix F**. Studies included to answer Key Question 1 were not formally evaluated for the strength of evidence, but key concepts of strength of evidence are discussed.

The strength of evidence was rated for Key Questions 2 using the four categories recommended in the AHRQ Methods Guide:[23,29] A "high" grade indicates high confidence that the estimate of effect lies close to the true effect for this outcome. The body of evidence has few or no deficiencies and the findings are stable (i.e., another study would not change the conclusions). A "moderate" grade indicates moderate confidence that the estimate of effect lies close to the true effect for this outcome. The body of evidence has some deficiencies and findings are likely to be stable, but some doubt remains. A "low" grade indicates low confidence that the estimate of effect lies close to the true effect for this outcome. The body of evidence has major or numerous deficiencies (or both) and additional evidence is needed before concluding either that the findings are stable or that the estimate of effect is close to the true effect. An "insufficient" grade indicates inability to estimate an effect, or no confidence in the estimate of effect for this outcome. No evidence is available or the body of evidence has unacceptable deficiencies, precluding reaching a conclusion.

Peer Review and Public Commentary

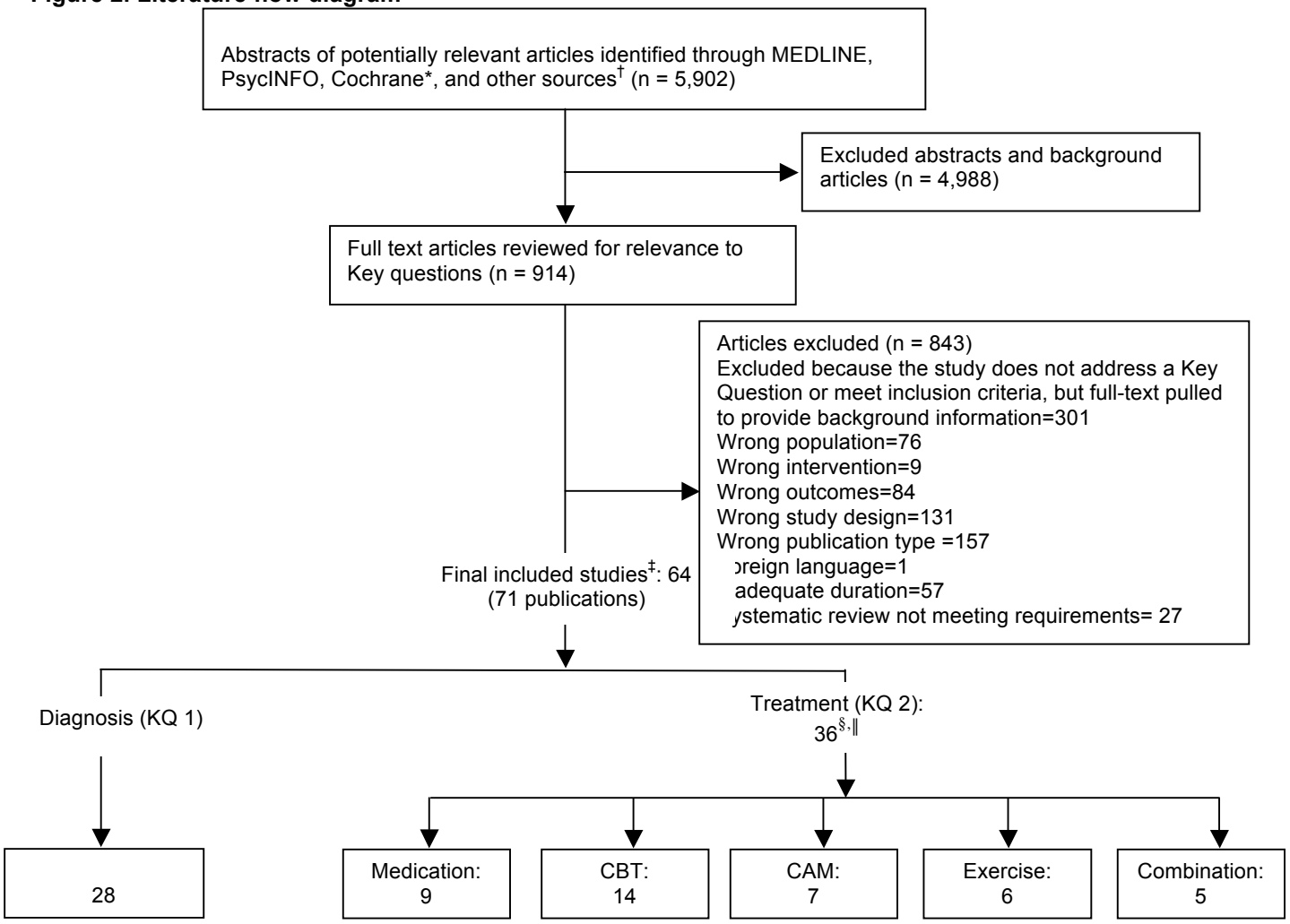
Experts in ME/CFS, individuals representing important stakeholder groups, and TEP members have been invited to provide external peer review of this systematic review. The AHRQ Task Order Officer and a designated Evidence-based Practice Center Associate Editor will also provide comments and editorial review. To obtain public comment, the draft report will be posted on the AHRQ web site for 4 weeks. A disposition of comments report detailing the authors' responses to the peer and public review comments will be made available after AHRQ posts the final systematic review on the public web site.

Results

Results of Literature Searches

Results of the literature search and selection process are summarized in the literature flow diagram (**Figure 2**). Database searches results in 5,902 potentially relevant articles. After dual review of abstracts and titles, 914 articles were selected for full-text review. After dual review of full text articles, 64 studies (in 71 publications) were included. Data extraction and quality assessment tables for included studies by key question are available in **Appendixes G** and **H**.

Figure 2. Literature flow diagram



CAM = complementary alternative medicine; CBT = cognitive behavioral therapy; KQ = key question.
*Cochrane databases include the Cochrane Central Register of Controlled Trials, Database of Abstracts of Reviews of Effects, Health Technology Assessment, National Health Sciences Economic Evaluation Database, and the Cochrane Database of Systematic Reviews.
†Identified from reference lists, hand searching, suggested by experts, etc.
‡Studies that provided data and contributed to the body of evidence were considered 'included'.

§ Studies may have more than one published article, this number indicates the number of unique studies included; there were a total of 43 articles included.
‖ Studies may have provided data for more than one treatment area.

Description of Included Studies

Of the 64 studies included in this review, 28 observational studies addressed Key Question 1 pertaining to aspects of diagnosis. Most were of fair-quality, enrolled predominantly female patients, had small sample sizes, and were conducted in the United States and Western Europe. Thirty-six randomized trials were included for Key Question 2, addressing the benefits and harms of interventions to treat ME/CFS (9 for medications, 14 for counseling and behavioral therapies, 7 for CAM, and 6 for exercise, including 5 comparing interventions). Most were of fair- or poor-quality, enrolled predominantly female patients from ME/CFS specialty clinics based on the CDC (Fukuda, 1994) or Oxford (Sharpe, 1991) case definition, had small sample sizes, and were conducted in the United States and Western Europe.

Key Question 1

What methods are available to clinicians to diagnose ME/CFS and how do the use of these methods vary by patient subgroups?

Key Points

- No studies using adequate sizes and spectrums of patients evaluated a diagnostic test for ME/CFS and no studies demonstrated an accurate and reliable method for identifying patients and subgroups of patients with ME/CFS.
- Several studies identified measures that appear to distinguish ME/CFS patients in individual studies, however none of these findings have been validated in other studies or in a second population.
- One good-quality study conducted in a broad but small spectrum of patients found that the artificial neural network test using a combination of 24 symptoms had good sensitivity, specificity, and accuracy. CFS symptoms with greatest single-item accuracy were acute onset of fatigue and sore throat. Further testing in populations with more diagnostic uncertainty is necessary to validate that these 24 symptoms can discriminate between patients with ME/CFS and those without the condition.

Detailed Synthesis

Eleven cross-sectional and longitudinal studies with comparison groups evaluated methods currently used to diagnose ME/CFS and provided data on discriminative value (ROC/AUC), sensitivity/specificity, or concordance of diagnoses (**Appendix G1**).[37-47] One study met criteria for good-quality,[38] nine for fair-quality,[39-47] and one for poor-quality[37] (**Appendix H1**). The studies were conducted in the United States[40,41,45-47] and western Europe,[37-39,42-44] were generally small (range: 25-798, with only one study enrolling >200 participants) and predominately enrolled women (43-100% female when reported). Major limitations of studies included fewer than 50 total subjects,[37,42-47] recruitment from specialty clinics,[38,45,46] lack of clear blinding to the reference standard result,[37-47] and compared cases with either healthy or non-fatigued controls.[37,39,42-46] Several studies used the same or very similar study populations to report on

different outcomes, recruiting from CFS self-help groups,[42-44] physician referral,[45,46] or a community sample outside Chicago.[40,41]

Overall the identified studies lacked robustness in evaluating diagnostic tests for ME/CFS. They were of small size and, with two exceptions, evaluated populations that consisted of ME/CFS cases and healthy controls rather than a broader spectrum of patients. Only one study used a population with overlapping symptoms and tested a strategy for diagnosis in both a derivation and a validation cohort.[38] While one tool, the artificial neural network, reported good sensitivity, specificity, and accuracy in a small derivation and validation cohorts,[38] it needs to be tested in broader populations with appropriate diagnostic uncertainty to further assess its diagnostic value. All other tests including serum markers and cardiopulmonary measures have not been adequately studied to determine their diagnostic usefulness in ME/CFS.

Biomarkers as Diagnostic Tests

All four of the studies of biomarkers were poor-quality and small (sample size range: 25-42); three utilized the same CFS self-help group population in Germany.[42-44] These studies evaluated the ability of serum parameters to identify ME/CFS versus healthy controls, and reported on the AUC for these measures. The tests included hypothalamic-pituitary axis testing (cortisol response to dexamethasone suppression test),[43] insulin tolerance testing and adrenocorticotropin hormone (ACTH), plasma and salivary cortisol responses to insulin injection,[42] pro-inflammatory cytokine response to standardized psychological stress,[44] and RNase L-isoforms.[37] Three of the small biomarker studies from the same group of investigators,[42-44] found that biochemical responses to stimuli were abnormal in the ME/CFS group compared with healthy controls. The morning plasma and salivary cortisol responses to low-dose overnight dexamethasone suppression testing were significantly lower in the ME/CFS group versus controls ($F=12.16$, $p=0.003$ for morning cortisol, $F=11.51$, $p=0.001$ for salivary free cortisol); this finding was consistent when comparing the logAUC (total) between groups.[43] The AUC of ACTH response to insulin tolerance testing was significantly associated with reported duration of symptoms ($F=4.92$, $p=0.03$); but there were no differences between ME/CFS patients and controls for plasma total and salivary free cortisol ($F=0.73$, $p=0.4$; $F=2.12$, $p=0.15$).[42] Response to stress was tested using the Trier Social Stress Test (TSST), a standardized psychological stress test, and found an inverted pro-inflammatory cytokine response for ME/CFS subjects compared with controls; ME/CFS subjects' levels of IL-6 and TNF-α decreased at 10 minutes and returned to normal by 60 minutes, whereas the IL-6 and TNF-α levels for controls increased at 10 minutes and returned to normal at 60 minutes (IL-6 $F=3.93$, $p=0.03$; TNF-α $F=4.64$, $p=0.02$).[44] ACTH response also varied between the groups, but cortisol did not (AUC for ACTH response curve $F=6.34$, $p=0.02$; AUC for plasma cortisol $F=0.1$, $p=0.91$; AUC for salivary cortisol $F=1.03$, $p=0.32$).[44] The fourth of these biochemical studies evaluated the sensitivity and specificity of RNase L levels in peripheral blood mononuclear cells for discrimination of ME/CFS subjects from controls; the ratio of RNase L isoforms at a cutoff of 0.4 had a sensitivity of 0.91 and specificity of 0.71; other thresholds results in lower sensitivity and specificity.[37] Although these tests were able to distinguish between healthy controls, without testing in a broader spectrum of patients including those with overlapping features, the usefulness remains uncertain.

Self-Reported Symptom Scales as Diagnostic Tests

Three small, fair-quality studies that reported on cortisol testing also reported the AUC values for the MFI-20, the Hospital Anxiety and Depression Scale (HADS), the Symptom

Checklist 90, Revised (SCL-90-R) and the Sickness Impact Profile 8-item (SIP-8) using essentially the same patient population and found that all measures were significantly different between ME/CFS cases and healthy controls (no medications, no current/lifetime psychiatric symptoms or disorders).[42-44] While AUCs were different, these studies do not further the diagnostic strategy for ME/CFS because of their methodological limitations: small size; case-control design; unclear recruitment methods, and unclear reporting of attrition and blinding (**Appendix H1)**. These results show that patients with ME/CFS have more depression, anxiety, and decreased functionality in several other domains; but because the comparison population consisted of healthy controls there is no evidence that these tests could adequately distinguish a ME/CFS population from another population of depressed, anxious, or medically ill patients. Overall, it is unclear whether these measures could diagnose ME/CFS if used by themselves (in the absence of the clinical criteria), because alone these measures do not satisfy the multiple symptom domains that currently comprise the syndrome of ME/CFS.

Two studies evaluated the ability of existing symptom scales to identify ME/CFS or to correlate with specific aspects of the diagnostic criteria such as disability or fatigue, in hopes of providing a more standard assessment tool for use in diagnosing ME/CFS. A fair-quality small study of 24 ME/CFS patients and 84 healthy controls, evaluated the SF-36, the CDC Symptom Inventory, and the MFI-20 for identifying ME/CFS subjects who met the disability criterion for the 2005 Reeves criteria.[48] The MFI-20 had reasonable sensitivity for the Reeves fatigue criteria but poor specificity (0.95 and 0.27 respectively); none of the AUCs for the MFI-20 were above 0.90.[40] In this study, the CDC Symptoms Inventory had poor sensitivity and specificity, as did the SF-36 subscales of physical functioning, role physical, social functioning, and role-emotional (none with AUC, sensitivity, or specificity above 0.90).[40]

In a subsequent paper, also fair-quality, the SF-36 was further evaluated using two different ME/CFS populations – 32 ME/CFS patients recruited from the community and 114 recruited from tertiary care – and 47 non-fatigued controls. Similar to the previous findings, none of the AUCs for the community-based ME/CFS population were above 0.90, whereas three AUCs for subscales of the SF-36 in the tertiary care CFS population were close to or above 0.90 (vitality, role-physical, and general health all had AUC of 0.91; social functioning had AUC of 0.87). Additional analysis focused on vitality, role-physical, and social functioning to determine cutoffs and assess whether the use of combinations of scales could identify ME/CFS subjects in both the community and the tertiary care samples as distinguished from healthy controls. They determined that meeting the cutoffs for two or more of these three subscales could be used to designate substantial reductions in function and to potentially distinguish those with ME/CFS from those without ME/CFS; for the community-based ME/CFS sample, sensitivity was 0.93 and specificity was 0.75; for the tertiary care sample, sensitivity was 0.96 and specificity was 0.75.[41] These researchers also used the MFI-20, the CDC Symptom Inventory, and the SF-36 to assess the sensitivity and specificity of the 2005 Reeves criteria for identifying ME/CFS in the community population compared with healthy controls; the AUC for Reeves criteria was 0.70 (sensitivity 0.65; specificity 0.76).[41] These studies do not appear to contribute to operationalizing the ME/CFS criteria given the inconsistencies in the results. The subscales of the SF-36 show promising results in a tertiary care – recruited population (the SF-36 scores for vitality, role-physical, and general health were above 0.90);[40] however, this was not true for the community-recruited ME/CFS patients, nor in a separate study.

Two studies created new assessment tools.[38,39] One good-quality study evaluated an appropriately broad spectrum of subjects (41 with systemic lupus erythematosis, 58 with

fibromyalgia, and 99 ME/CFS patients) and randomly assigned them to either a derivation or validation cohort.[38] A new tool was developed by administering prospectively defined criteria via questionnaire; each symptom was assessed for sensitivity and specificity and the symptoms with the best sensitivity and specificity were elected to contribute to the new criteria. Four methods for classification of ME/CFS were tested using the derivation cohort, and for each algorithm sensitivity, specificity, and accuracy were determined using the validation cohort. One of the four strategies that included 24 symptoms, the artificial neural network, had the best results (sensitivity 0.95; specificity 0.85 and accuracy 0.90).[38] Although the artificial neural network test appears to discriminate well, its generalizability is limited as it has not been reproduced or studied in a larger population. One other large (n=368 ME/CFS and 430 controls) and fair-quality study tested the Schedule of Fatigue and Angina for CFS scale (SOFA-CFS) in ME/CFS patients and healthy controls using latent class analysis and demonstrated good sensitivity and specificity (sensitivity 0.81, specificity 0.98).[39] Inherent risk of bias exists however when a test has not been validated in a population other than that in which it was derived. Such studies may exaggerate the predictive ability of models and, furthermore, may not be broadly applicable to populations of interest. Furthermore, the SOFA-CFS trial compared known ME/CFS patients with healthy controls rather than a spectrum of patients leaving uncertainty in the validity of this test.

Exercise Testing as a Diagnostic Test

Exercise testing was evaluated in two fair-quality studies of the same population.[45,46] The first of these studies demonstrated that cardiopulmonary exercise test (CPET) capacity was significantly different between ME/CFS subjects and non-disabled sedentary controls. SF-36 and MFI-20 were then tested to determine whether these two scales could distinguish those who would fail to recover from CPET within 1 day. The AUC analysis demonstrated that SF-36 subscales of physical function, role-physical, bodily pain, general health, vitality, and social functioning were significant for failure to recover at 1 day; and the subscales role-emotional, vitality, and bodily pain were significant for failure to recover at 1 week.[45] A separate study tested whether individual symptoms could identify ME/CFS subjects versus controls and found fatigue, neuroendocrine dysfunction, immune dysfunction, pain, and sleep disturbance all had significant AUC.[46] These studies are limited by small size and case-control design and precludes any valid conclusion about the utility of SF-36 or MFI-20 for prediction of failure to recover at 1 day or 1 week.

A fair-quality study used cluster analysis to identify coping strategies for ME/CFS patients, and was able to determine standardized discriminant function and structure coefficients for the three clusters.[47] One function separated the clusters and was significant ($F=3.31$, $p=0.01$) and accounted for 10 percent of the variables between groups ($R_c=0.32$). Adaptive coping accounted for 56 percent of the variance explained by the function ($R_s=0.75$) and less adaptive coping accounted for 25 percent ($R_s=0.50$).

Key Question 1a

What are widely accepted diagnostic methods and what conditions are required to be ruled out or excluded before assigning a diagnosis of ME/CFS?

Key Points

- Eight different case definitions have been used to identify a population of people with ME/CFS; all include a set of clinical criteria and are applied by clinicians. Despite being developed as consensus guidelines and with endorsement of national groups, none of these published diagnostic strategies have proved superior in distinguishing ME/CFS from other conditions that may also present with fatigue.
- Most ME/CFS case definitions require that other conditions be excluded prior to assigning a diagnosis of ME/CFS, however no studies compared strategies for ruling out alternative diagnoses.

Detailed Synthesis

The published case definitions have evolved since the first set of clinical criteria were published by the CDC in 1988 (**Table 1** below and **Appendix I**).[6] These case definitions address the diagnostic work-up that is required prior to diagnosing CFS. In general, prior to diagnosing ME/CFS, other explanations for fatigue are to be ruled out. Recommendations for work-up are included with the published case definitions, but no studies specifically evaluated diagnostic methods, or compared strategies for excluding other diagnoses prior to assigning a diagnosis of ME/CFS. Additionally, no one case definition is universally accepted.

Table 1. Case definitions and criteria

Name of Case Definition	Included Criteria
CDC – CFS Holmes, 1988[6]	Requires each of the following: 1) ≥6 months of persistent or relapsing, debilitating fatigue not resolved with bed rest 2) 8 of 11 minor symptoms: fever or chills, sore throat, lymph node pain, muscle weakness, muscle pain, PEM, headaches of a new or different type, migratory arthralgia, neuropsychiatric complaints, sleep disturbance, and a sudden onset of symptoms 3) ≥50% impairment of daily functioning as compared with premorbid levels
Oxford - CFS Sharpe, et al. 1991[49]	Requires each of the following: 1) Fatigue as principal symptom 2) Definite onset of syndrome (not lifelong) 3) Syndrome must be severe, disabling have an effect on physical and mental (cognitive) functioning; 4) Present for >6 months, or >50% of the time 5) May include other symptoms: myalgias, mood, and sleep disturbance
London – ME/CFS Dowsett, 1994[50]	Describes a spectrum of disease in which patient presents with severe idiopathic chronic fatigue. Includes CFS, ME, and post viral fatigue syndrome. Suggests conditions to rule out and categories of symptoms that are frequently present.

Name of Case Definition	Included Criteria
CDC - CFS Fukuda *et al.*, 1994[3]	Evaluate cases of prolonged or chronic fatigue with history/physical examination, mental status evaluation (psychiatric, psychologic and/or neurologic exams as appropriate) and lab tests to screen (CBC, ESR, ALT, TP, albumin, globulin, alkaline phosphate, calcium, phosphate, glucose, bun, electrolytes, creatinine, TSH, and urine analysis) and others as clinically indicated to exclude other conditions. Exclude cases where another cause of chronic fatigue is found. Classified as CFS if both: 1) Fatigue persists or relapses ≥6 months 2) ≥4of the following symptoms concurrently present for ≥6 months: impaired memory or concentration, sore throat, tender cervical or axillary lymph nodes, muscle pain, multi-joint pain, new headaches, unrefreshing sleep, PEM
Canadian – ME/CFS Carruthers *et al.*, 2003[1]	Required to meet the criteria for fatigue (significant degree of new onset, unexplained, persistent, or recurrent physical and mental fatigue that substantially reduces activity level), PEM (inappropriate loss of physical and mental stamina, rapid muscular and cognitive fatigability, PEM and/or fatigue and/or pain and a tendency for other associated symptoms within the patient's cluster of symptoms to worsen; pathologically slow recovery period - usually ≥24 hours) and/or fatigue, sleep dysfunction (unrefreshed sleep or sleep quantity or rhythm disturbances such as reversed or chaotic diurnal sleep rhythms), and pain (significant degree of myalgia; may be experienced in muscles and/or joints, and is often widespread and migratory in nature; often significant headaches of new type, pattern or severity.) 1) ≥2 neurological/cognitive manifestations 2) ≥1 symptoms from 2 of the categories of autonomic, neuroendocrine and immune manifestations. 3) Illness lasting ≥6 months, usual with distinct onset but may be gradual
CDC – CFS Reeves *et al.*, 2005[48]	Surveillance criteria (used self-report on questionnaire using items that correlated with Fukuda, 1994 case definition): 1) Fatigue (must satisfy all): - lasting >6 months - not relieved by rest (by answering "a little or not at all" to the question "is your fatigue relieved by rest?) - causing substantial reduction in occupational, educational, social, or recreational activities (by answering "a lot" to "Does fatigue interfere with...") 2) Presence of 4 of 8 case-defining symptoms (by answering "all of the time or most of the time" to questions about symptoms, e.g., “during the past month how often have you had a sore throat?") Standardized clinically empirical criteria: functional impairment, severe fatigue and reporting ≥4 symptoms and scoring ≥25 on the Symptom Inventory Case Definition Subscale.
Revised Canadian – ME/CFS Jason *et al.*, 2010[51]	6 months of persistent or recurring chronic fatigue that is not lifelong and results in substantial reductions in previous levels of function and includes concurrent occurrence of: post-exertional malaise and/or post-exertional fatigue; unrefreshing sleep or disturbance of sleep quantity/rhythm; pain - widespread and migratory; ≥2 cognitive manifestations; ≥1 symptoms from 2 of the categories - autonomic, neuroendocrine, immune; and excluding active medical conditions that may explain the presence of fatigue.
International Consensus Statement - ME Carruthers *et al.,* 2011[2]	1) Postexertionalneuroimmune exhaustion: cardinal 2) Neurological impairments: ≥ 1 from 3 of the 4 symptom categories (neurocognitive, pain, sleep, neurosensory/motor) 3) Immune, gastrointestinal and genitourinary impairments: ≥1 symptom from ≥3 categories 4) Energy production⁄transportation impairments: at least one (cardiovascular, pulmonary, thermostatic, temperature)

ALT= alanine amino transferase; CBC= complete blood count; CDC= Centers for Disease Control and Prevention; CFS= chronic fatigue syndrome; ESR= erythrocyte sedimentation rate; ME= myalgic encephalomyelitis; TP= total protein; TSH= thyroid stimulating hormone.

Key Question 1b

What is the accuracy and concordance of diagnostic methods?

Key Points

- Evaluation of accuracy of a diagnostic test generally requires an accepted diagnostic reference (gold) standard. Diagnostic studies of ME/CFS are limited by the lack of an accepted reference standard. No studies evaluated the accuracy of current diagnostic methods.
- Concordance was qualitatively assessed in nine studies, six reporting variations in symptom prevalence in populations that were defined by different case definitions; none have proved superior in distinguishing ME/CFS from other conditions that may also present with fatigue.
- Populations identified by criteria labeled as ME or ME/CFS criteria had more severe symptoms or more functional impairment than those labeled as CFS criteria.

Detailed Synthesis

No studies evaluated the accuracy of current diagnostic methods for ME/CFS, given that evaluation of accuracy of a diagnostic test generally requires an accepted diagnostic reference (gold) standard and that no such accepted reference standard exists for ME/CFS. Nine studies of diagnostic methods evaluated the concordance of different diagnostic criteria (**Appendix G2**).[4,5,52-58] These were primarily observational cohort studies and were descriptive, therefore they were not amenable to quality rating. Differences reported below are those that were statistically significant between groups.

Six studies evaluated various case definitions and demonstrated that symptom reporting varied between populations defined by different sets of criteria. In general, populations labeled as ME or ME/CFS criteria were more symptomatic and impaired than those labeled as CFS criteria alone (**Table 1** and **Appendix I**).[4,5,53,54,57,58] Three studies comparing CFS patients with healthy controls found differences in symptom reporting. The Fatigue Impact Scale (FIS), Chalder Fatigue Scale, HADS depression subscale (HADS-D), and certain SF-36 subscales or combinations of SF-36 variables with Zung Depression Scale may help to identify subgroups of CFS patients, but will need to be evaluated in a broad spectrum of patients with diagnostic uncertainty to determine their ability to differentiate between conditions and/or identify clinical subgroups of patients.[52,55,56] For more detail on the scales used in these studies see **Appendix J**.

Studies Comparing Case Definitions for ME/CFS

Six studies evaluated one or more populations using two or more case definitions for ME/CFS and attempted to demonstrate differences between diagnostic criteria.[4,5,53,54,57,58] Sore throat and lymph node pain were more common in 14 subjects who met the Holmes, 1988 criteria compared with two other groups (CDC [Fukuda, 1994] criteria, n=18 and fatigued patients due to psychiatric illness, n=33).[54] This study also compared the SF-36 among these groups and found no differences (bodily pain varied between 1988 CFS group and psychiatric illness group but not between 1994 CFS group and psychiatric illness group; general health

varied between the 1998 and 1994 CFS groups, but not between either fatigued or the psychiatric illness group; physical health composit score varied between 1988 CFS group and the psychiatric illness group, but not between 1994 CFS group and the psychiatric illness group; mental health composite score had no differences between any groups; self-reported degree of impairment varied between 1994 and the psychiatric illness group but not between 1988 and the psychiatric illness group. A similar study compared symptom prevalence for CFS identified by CDC (Fukuda, 1994) criteria with ME/CFS identified by Canadian (Carruthers, 2003) criteria using data from three populations, comparing scores on the DePaul Symptom Questionnaire and the SF-36.[4] The SF-36 scores indicated less impairment for the CFS group compared with the ME/CFS group in all three populations on the subscales of physical functioning and bodily pain. Symptom reporting indicated less impairment in the CFS group compared with the ME/CFS group in the majority of PEM, pain, autonomic, and immune symptom subcategories; responses to other symptom subcategories were either partial (4/13 for all 3 populations in the case of neurocognitive symptoms for instance) or inconsistent across the populations (only 1/6 sleep symptoms was lower for CFS vs. ME/CFS in all 3 populations, whereas 3/6 sleep symptoms were lower in 2 of the populations).

Another study compared 74 patients labeled as CFS defined by CDC (Fukuda, 1994) criteria with 39 patients labeled as ME defined by the international consensus criteria using the SF-36 and symptom scales.[53] In this study, SF-36 subscale scores indicated less impairment among the CFS group versus the ME group on the physical functioning, bodily pain, vitality, and social functioning subscales. Symptom ratings also indicated less impairment among the CFS group compared with the ME group for PEM, neurological, and pain symptoms.

Using a similar population, another study compared the Canadian ME/CFS definition to the CDC CFS (Fukuda, 1994) criteria and an ME definition created from Ramsey 1988 and others, based on the cardinal features of ME (acute onset plus PEM, neurological manifestations, and autonomic manifestations).[5] Of the 114 people who met the CDC (Fukuda, 1994) criteria for CFS, 56 were also classified as ME/CFS and 27 as ME. There were significant differences among these groups in multiple symptoms; symptom reporting was lower for the group who met CDC (Fukuda, 1994) criteria (but not ME) versus those who met the Canadian ME/CFS criteria, and lower for those who met the CDC (Fukuda, 1994) criteria (but not the ME/CFS criteria) versus those who met the ME criteria (defined in this study and based on prior definitions and cardinal features). The ME/CFS group had higher psychiatric comorbidity rates compared with the CFS group. Objective measures of heart rate, cognitive function (trail making tests), and the Kroenke 13 symptom inventory were also compared across groups, demonstrating that ME and ME/CFS groups had higher heart rates lying down and 2 and 10 minutes after standing compared with the CFS-only group, longer times on the trail making tests, and higher scores on the Kroenke symptom and psychiatry comorbidity scale.

Another study compared 41 subjects meeting CDC CFS (Fukuda, 1994) criteria with 26 subjects meeting London/National Task Force 1994 ME criteria using the SF-36, MFI-20, Karnofsky Performance Scale (KPS), and exercise testing. ME subjects had lower functioning than CFS subjects on the role-emotional and mental health subscales of the SF-36 (others were not significantly different). The role-emotional subscale correctly classified 60 percent of the cases (73% of ME cases and 51% of the CFS cases). General fatigue scores were higher for CFS versus ME; certain physical parameters were higher for ME subjects versus CFS.[58]

Using a unique approach, another study evaluated whether symptoms vary for younger versus older CFS patients. They studied 50 CFS subjects, (CDC [Fukuda, 1994] criteria) 25 over

age 50 years matched by sex and duration of CFS diagnosis with 25 subjects ages 16 to 29 years.[57] Older CFS patients had higher FIS scores, higher Chalder Fatigue scores, higher HADS-D scores, lower functioning by SF-36, lower self-efficacy, and several autonomic and hemodynamic differences compared with younger CFS patients. The two groups did not differ on the Cognitive Failures Questionnaire, HADS total, HADS anxiety subscale (HADS-A), pain rating, Epworth Sleepiness Scale, and Orthostatic Grading Scale.

In summary, patients being diagnosed using case definitions labeled as ME or ME/CFS reported greater symptoms and had more impairment. The results suggest that the CFS criteria captures a broader population, and that ME or ME/CFS criteria identify subsets with greater severity of symptoms from among the CFS group. Differences may exist between younger and older patients with CFS but this has only been studied in one small study so is inconclusive.

Studies Comparing Symptoms Among ME/CFS and Non-ME/CFS Populations

Three studies attempted to compare ME/CFS subjects (CDC [Fukuda, 1994] criteria) with other non-ME/CFS groups (either healthy controls or controls with other disease states). One used variables from the SF-36 and Zung Self-Rating Depression Scale among 51 women with CFS, 55 with idiopathic chronic fatigue defined as chronic fatigue not meeting criteria for CFS, and 53 non-fatigued controls matched to the CFS subjects.[52] In this study, latent class analysis empirically derived a solution that was comparable with the established definitions, in essence validating the ability of these criteria for distinguishing patient groups that differ.

A second study evaluated psychiatric symptoms in 98 consecutive patients presenting to an academic medical center's chronic fatigue clinic, 19 of whom met CDC (Fukuda, 1994) criteria for CFS (79 had fatigue but did not meet criteria) and a comparison group of 31 subjects with rheumatoid arthritis (RA). General Health Questionnaire scores were highest for the chronic fatigue group and lowest for the RA group; the SF-36 role function scores indicated lowest impairment in RA group and highest in the CFS group; SF-36 mental function was best in the RA group and lowest in the chronic fatigue group; SF-36 health perception was highest in the RA group and lowest in the CFS group; no differences in the other SF-36 subscales or in the Modified Symptoms Perception Questionnaire or the Pennebaker Inventory of Limbic Languidness.[55]

A final study compared functional status and well-being of 223 patients who met the CDC CFS (Holmes, 1988) criteria with both a population-based control sample and a group with various chronic diseases using the SF-36.[56] CFS patients had lower functioning than the general population on all SF-36 subscales, and lower functioning than almost all disease groups on most subscales: the exceptions were that the CFS group did not differ from the group of 25 multiple sclerosis patients in terms of physical functioning, vitality, and role-emotional, nor did the CFS group differ from the congestion heart failure group on the role-emotional subscale.

Based on these three studies, symptom reporting varies between CFS patients and other populations but the utility of these symptom-based scales in differentiating patients with diagnostic uncertainty remains inconclusive.

Key Question 1c

What harms are associated with diagnosing ME/CFS?

Key Points

- Five studies found that patients with ME/CFS feel stigmatized by their diagnosis in terms of financial stability, work opportunities, perceived judgments on their character, social isolation, and interactions with the health care system.
- Prejudice and stereotypes within the medical profession have been identified in two studies; medical trainees and mental health practitioners make judgments about a patient's condition based on the name it carries (ME, CFS, or other) and what treatment is being given.
- There is a substantial burden of misdiagnosis among the CFS population.

Detailed Synthesis

Harms of the diagnostic process or the diagnosis of ME/CFS were evaluated in eight studies using primarily descriptive methods that were not amenable to quality rating (**Appendix G3**).[59-66] Two studies are based on surveys[61,64] and three on qualitative interviews[60,63,65] to assess patients experiences and understandings of their disease. One study performed thorough psychiatric evaluation to identify the frequency of missed psychiatric disease in CFS.[62] Two studies randomized participants to various disease names and gave identical case descriptions in order to test the effect of the disease name on perceptions by medical trainees and undergraduate students.[59,66] CFS patients feel stigmatized by their diagnosis in terms of worsened financial stability, fewer work opportunities, perceived judgments on their character, social isolation, and interactions with the health care system. Compounding this is a substantial burden of misdiagnosis among this patient population. Objectively measured and identified prejudice and stereotypes exist within the medical profession; medical trainees and mental health practitioners make judgments about a patient's condition based on the name it carries (ME, CFS, or other) and what treatment is being given.

The five studies that used either survey or interview methods to assess harms found that CFS patients experience social stigma as a result of their disease. These include decrease in financial stability (financial losses and lower standard of living, new job that required fewer skills, or pay cut), decrease in social life and loss of friends and feeling estranged, feeling like they needed to conceal their symptoms, and difficult interactions with the medical profession (stereotypes perpetuated and doctors having decided before meeting them that they had a psychological diagnosis), and feeling like their moral character was questioned.[60,61,63-65]

One of these studies performed a prospective evaluation of the frequency of misdiagnosis in patients with CFS by studying 68 patients who met the Oxford (Sharpe, 1991) criteria for CFS and conducting standardized structured interviews with a consultant psychiatrist, along with a full medical, psychiatric, family, and personal history. Of 68 patients evaluated, 31 (46%) reported having been given a psychiatric diagnosis (2/3 of them had been incorrectly diagnosed).[62] Specifically, 21 patients had been given a psychiatric diagnosis when one did not

exist, and 13 patients who had never been given a psychiatric diagnosis actually had a treatable psychiatric condition in addition to CFS.[52]

Two publications describe a study of undergraduate students (n=105) and medical trainees (n=141) who were randomized to being told that the diagnosis for a patient case presentation (identical among all groups) was either CFS, ME, or Florence Nightingale Disease.[59,66] Medical trainees' perceptions of diagnostic accuracy, physiological etiology, and prognosis varied between groups; CFS label was considered most accurate, while the ME label carried worse prognosis. Mental health practitioners were randomized to being told that an identical CFS patient was getting one of three treatments. The assigned treatment appeared to influence subsequent attributions of the patient's disease. Specifically, practitioners who were told that the patient was getting an intravenous immune modulator as the treatment were more likely to think that the patient was correctly diagnosed as having CFS and was more disabled ($p<0.05$ for both).[59]

Key Question 2

What are the (a) benefits and (b) harms of therapeutic interventions for patients with ME/CFS and how do they vary by patient subgroups?

Key Points

- Thirty-six trials provided evidence of benefits and harms of therapeutic interventions; all were small, most of fair-quality, and comparison limited by variability in scales used to measure outcomes.

Medications

- Nine trials met inclusion criteria for medical treatment of ME/CFS, although none of the medications have been approved by the FDA for this indication.
- Two fair-quality trials of rintatolimod, an immune modulator, enrolling severely debilitated participants found improvement in measures of exercise performance (low strength evidence). Improvement in other measures of function and reduction of other medications for relief of CFS symptoms was also found in one of the studies but provided insufficient evidence.
- All other evidence was insufficient due to small single studies with methodological limitations.
- A small fair-quality trial of valganciclovir enrolled participants with suspected viral onset of ME/CFS and elevated antibody titers and reported improved fatigue compared with placebo based on one scale, but no differences for other measures.
- Small single trials of Isoprinosine, hydrocortisone, immunoglobulin G, and fluoxetine did not show significant improvement compared with placebo. Differences were also not found in a larger dose-ranging trial of galantamine.
- Harms of medications included suppression of adrenal glucocorticoid responsiveness, increased appetite, weight gain, and difficulty sleeping with hydrocortisone; flu-like syndrome, chills, vasodilatation, dyspnea, and dry skin with rintatolimod; and headaches with immunoglobulin G. Withdrawals due to harms were greater with fluoxetine than placebo.

Counseling and Behavior Therapies

- When considering all studies comparing any type of counseling with no treatment, support, relaxation, or adaptive pacing there is moderate strength of evidence that counseling improves fatigue (9/12 trials showed positive effect), low strength of evidence for measures of functioning (5/12 trials showed positive effect; 2/12 showed mixed results on different measures), low strength of evidence for quality of life (2/5 trials showed positive effect; 1/5 showed mixed results on different measures), and moderate strength of evidence for global improvement (3/3 trials showed positive effect).
- Low strength evidence suggests that at followup, patients receiving counseling had better SF-36 physical functioning subscale scores than control patients, based on a pooled analysis of eight trials; weighted mean difference in score of 7.73 (95% CI, 3.58 to 11.87).
- There is low strength of evidence from a small fair-quality trial that face-to-face counseling is similar to telephone counseling with both improving function, employment measures, and global change.
- Harms of counseling and behavioral therapies were poorly reported but there is low strength of evidence that counseling is not associated with harms based on one moderate-sized trial.

Complementary and Alternative Medicine

- All outcomes studied have insufficient evidence due to small single studies with methodological limitations.
- Two small fair-quality trials of CAM interventions (one with homeopathy and one with L-carnitine preparations) found improvement in some measure of fatigue and/or function but other trials did not.
- One study found that being aware that one is not receiving distant healing resulted in smaller improvements in function based on one good-quality study.
- All other trials of CAM interventions found no significant improvements compared with placebo, usual care, or an alternative CAM approach.
- Adherence was low in one trial of a low sugar/low yeast diet but otherwise adherence and harms were not well reported.

Exercise Therapy

- Graded exercise treatment (GET) was superior to control groups in measures of fatigue (low strength) , function (moderate strength), and clinical global impression of change (moderate strength) based on one-good quality and three fair-quality randomized trials.
- Although single small studies found qigong exercise provided improvement in measures of fatigue and that home orthostatic training was similar to usual care or sham orthostatic training, this evidence was insufficient.
- Harms were not well reported overall, and evidence is insufficient. Patients receiving GET reported more harms compared with cognitive behavioral therapy (CBT), adaptive pacing, or usual care in one good-quality trial and almost half of patients assigned to physiological exercise testing (10/25) refused to repeat testing at followup over concern

for harm. Dropout rates were greater with exercise (25/68, 37%) than fluoxetine or placebo (15/69, 22%).

Combination Therapy and Head-to-Head Comparisons

- Low strength evidence suggests that GET and CBT had similar results on measures of fatigue and function based on three trials.
- Evidence on the comparison of GET and fluoxetine is insufficient because there was only one small study with methodological flaws. This study found GET superior to fluoxetine on measures of fatigue and function. There is low strength of evidence that CBT or GET are not associated with an increase in serious harms.

Detailed Synthesis

Thirty-six trials of interventions for patients with ME/CFS in 43 publications met inclusion criteria; nine of medications, 14 of counseling and behavior therapies, seven of CAM interventions, four of exercise programs, and four of combinations of these interventions (**Appendix H4**). Seven were rated good-quality, while 24 were rated fair- and five poor-quality (**Appendix I2**).

Trials enrolled from 22 to 641 patients with ME/CFS and most (26/36, 72%) used the CDC (Fukuda, 1994) diagnostic criteria. Outcomes measures included the SF-36 physical functioning subscale, Medical Outcome Study Short Form (MOS-SF), Checklist of Individual Strength (CIS), Profile of Mood States (POMS) fatigue subscale, KPS, and SIP-8 scale to measure overall function; MFI-20, Chalder Fatigue Scale, Krupp Fatigue Severity Scale (FSS), FIS, and Visual Analogue Scale (VAS) to measure fatigue; Quality of Life Inventory (QOLI), Quality of Life Index (QLI), Quality of Life Scale (QLS), EuroQol Scale, Global Wellness Scale, Short Form 12-item Health Survey (SF-12), and Fibromyalgia Impact Questionnaire (FIQ) to measure quality of life; Clinical Global Impression Change (CGI) scales to measure improvement over time; and the Work and Social Adjustment Scale to measure impairment in work. These are described in **Appendix J**.

Medications

Nine randomized trials provided evidence for the medical treatment of ME/CFS, including placebo-controlled trials of galantamine,[67] hydrocortisone,[68] hydrocortisone plus fludrocortisone,[69] immunoglobulin G,[70] valganciclovir,[71] rintatolimod,[72,73] isopinosine,[74] and fluoxetine.[75] (**Table 2** below; **Appendix G4**). None of these medications have been approved by the FDA for this indication.

Eight trials met criteria for fair-quality,[67-73,75] and one for poor (**Appendix H2**).[74] Major limitations of studies include enrolling fewer than 20 subjects in an arm,[70,71,74] high loss to followup,[67,75] lack of intention-to-treat analysis of outcomes,[69,72] lack of reporting between-group comparisons for key outcomes,[73] unclear randomization process,[74] and lack of blinding.[74] Most trials were either funded by pharmaceutical companies (fully or in part)[67,70-73] or the funding source was not reported.[68,69]

Most trials were designed to treat the potential underlying pathology of ME/CFS. All but two trials[74,75] enrolled participants in the United States; only three enrolled more than 100 participants;[67,73,75] and three were multi-center.[67,72,73] Participants were predominantly women, and their mean ages ranged from 32 to 50 years. Although most participants were white, many trials did not report race or ethnicity. Most trials used the CDC (Fukuda, 1994) criteria in their

inclusion criteria except for one study predating it that used the CDC (Holmes, 1988) criteria,[70] and one trial that used the Oxford criteria.[75] The duration of illness varied widely with some trials enrolling participants with durations under 3 years[68,69,74,75] while others were 10 years or more.[71,73]

Outcome measures of fatigue differed between trials, precluding direct comparisons. These included the CGI scale, Chalder Fatigue Scale, POMS (fatigue and vigor subscales only), VAS (degree of fatigue; abbreviated fatigue questionnaire), FSS, fatigue scale specific to the trial, hours of rest per day, Symptom Severity Scale (fatigue, prolonged post-exertion fatigue), CPET tolerance, exercise duration and work, MFI-20, and CDC CFS Symptom Inventory. Additional measures of function and quality of life were also used as outcomes.

Antiviral and/or immune modulators resulted in improvement in fatigue symptoms in patients with ME/CFS, while trials of intravenous rintatolimod reported improved measures of function and reduced use of other medications for relief of CSF symptoms. Trials of galantamine, hydrocortisone, immunoglobulin G, and fluoxetine indicated no significant improvement compared with placebo.

A trial of valganciclovir, an antiviral agent, enrolled 30 participants with suspected viral onset of ME/CFS and elevated antibody titers.[71] The treatment group received oral valganciclovir 900 mg twice daily for 21 days, then 900 mg once daily for total of 6 months. Participants were followed for another 6 months, and unblinding and outcomes were measured at 9 months. Differences were statistically significant from placebo for scores on the FSS (-0.06 for valganciclovir vs. 0.02 for placebo; p=0.006), but not the MFI-20, CDC CFS Symptom Inventory, or self-reported physical function. Attrition was 9 percent and adherence 91 percent. No harms were reported for either group**.**

Three trials compared drugs with immunomodulatory effects with placebo, including trials of intravenous rintatolimod[72,73] and oral Isoprinosine, both not currently FDA approved for any indication in the United States.[74] In an early trial of rintatolimod, 92 severely debilitated patients (KPS scores of 20-60) were randomized to rintatolimod 200 mg twice weekly for 4 weeks, then 400 mg twice weekly for a total of 24 weeks or placebo.[72] The median percentage changes from baseline to week 24 were statistically significantly different between groups for exercise duration (10.3 for rintatolimod vs. 2.1 for placebo; p=0.007), exercise work (11.8 for rintatolimod vs. 5.8 for placebo; p=0.011), activities of daily living (23.1 for rintatolimod vs. 14.1 for placebo; p=0.034), and KPS (20 for rintatolimod vs. 0 for placebo; p=0.023). Attrition was 9 percent and adherence 91 percent, and harms did not differ between groups. This trial was limited by lack of intention-to-treat analysis. A second trial randomized 240 participants (KPS scores of 40-60) to rintatolimod 400 mg twice weekly for 40 weeks or placebo.[73] The mean percentage change in CPET tolerance from baseline to week 40, the primary outcome, was greater for the treatment versus placebo group (37% vs. 15%; p=0.047). Although other performance scores were measured, they were not compared between groups (KPS, activities of daily living, SF-36 vitality and general health perception subscales). More participants in the treatment group reported decreased use of medications for relief of CFS symptoms (68% vs. 55%; p=0.048). Attrition was 19 percent and adherence 83 percent. Flu-like syndrome, chills, vasodilatation, and dyspnea were more frequent in the treatment group ($p<0.05$). A single-blinded trial of Isoprinosine randomized 10 patients to treatment and 6 to placebo.[74] The treatment group received 3 g/day of Isoprinosine in divided doses for 12 weeks that varied over time. Mean changes in KPS scores from baseline did not differ between groups (p=0.93).

To evaluate the efficacy of galantamine, an acetyl-cholinesterase inhibitor, participants from 35 clinical centers in the United Kingdom, western Europe, and United States were randomized to oral galantamine at various doses (7.5, 15, 22.5, or 30 mg/day) or placebo for 16 weeks (8 weeks at full dose).[67] Outcome measures indicated no statistically significant differences or dose effect between groups for the primary outcome of global improvement (CGI scale), or secondary outcomes of fatigue (Chalder Fatigue Scale), and quality of life (FIQ). The overall withdrawal rate was 23 percent and attrition rate 30 percent, but rates were highest among groups given galantamine doses of 15 mg or more per day. Overall, 90 percent reported harms, with depression, nausea, and headache most common in both groups. Two percent of the galantamine participants experienced serious events, but none was attributed to the study drug.

Two trials evaluated corticosteroids versus placebo, including a trial randomizing 70 participants to oral hydrocortisone (20-30 mg every am and 5 mg every pm) or placebo for 12 weeks,[68] and a crossover trial of 100 participants using hydrocortisone (5 mg/day) plus 9-alpha fludrocortisone (50 µg/day) for 12 weeks.[69] Neither study reported statistically significant differences in outcomes between treatment and placebo groups for fatigue (POMS; VAS), quality of life (Global Wellness scale; VAS), or function (activity scale; SF-36). Attrition rates were 10 percent[68] and 20 percent.[69] Harms that significantly differed between treatment and placebo groups included suppression of adrenal glucocorticoid responsiveness (12 vs. 0; $p<0.001$); increased appetite (17 vs. 8; $p=0.02$); weight gain (19 vs. 8; $p=0.006$); and difficulty sleeping (17 vs. 8; $p=0.02$).[68]

Intravenous immunoglobulin G (1 gm/kg) versus placebo (1% albumen solution) given once every 30 days for 6 months was evaluated in a trial of 30 participants.[70] While measures of fatigue, prolonged post-exertion fatigue (Symptom Severity Scale), and physical function (MOS-SF) were not statistically significantly different between groups, social function (MOS-SF) improved for the placebo group ($p<0.05$). Overall, attrition was 7 percent, and 20 percent experienced harms including 93 percent of treatment and 60 percent of placebo participants reporting headaches ($p=0.03$).

Fluoxetine was compared with placebo in a 6-month, 4-arm fair-quality trial that also included a GET group which will be described separately below.[75] Differences between fluoxetine and placebo were not statistically or clinically significant for fatigue (Chalder Fatigue Scale), functional capacity measured as the amount of oxygen consumed in the final minute of exercise per kg of body weight, or rates of non-fatigue (Chalder Fatigue score of <4). Attrition was higher with fluoxetine than placebo (32% vs. 17%) and adherence was not reported. Withdrawals due to medication side effects were greater with fluoxetine (9/68, 13%) than placebo (2/68, 3%) although there were no differences in total withdrawals.

In summary, there is low strength evidence that rintatolimod improves measures of exercise performance. Although a trial of valganciclovir indicated improvement of fatigue among patients with suspected viral onset of ME/CFS and elevated antibody titers, and one trial of intravenous rintatolimod reported improved measures of KPS, activities of daily living, and reduced use of other medications for relief of CSF symptoms, the strength of evidence for these outcomes and all other medications is insufficient because each medication was evaluated by only one small trial with important methodological limitations (**Appendix K**) and few differences were found between treatment and placebo groups.

Table 2. Trials of medications for ME/CFS

Author, year Study type N Quality	Case definition	Duration/ followup	Interventions	Overall Effect: Treatment compared with placebo
Blacker, *et al.,* 2004[67] N=423 Fair	CDC (Fukuda, 1994) criteria	4 months (16 weeks, 8 weeks at full dose)	A. Galantamine 2.5 mg TID B. Galantamine 5 mg TID C. Galantamine 7.5 mg TID D. Galantamine 10 mg TID E. Placebo	**Fatigue:** Chalder Fatigue Scale (mean change from baseline) Physical scores: NS Mental scores: NS **Quality of life:** FIQ (mean change from baseline): NS Global Well Being (composite): NS **Other:** % Improvement on modified CGI: NS
Blockmans, *et al.,* 2003[69] N=80 Fair	CDC (Fukuda, 1994) criteria	3 month treatment; 3 month placebo crossover	A. Hydrocortisone 5 mg/day + 9-alpha fludrocortisone 50 µg/day B. Placebo	**Fatigue outcomes**: VAS degree of fatigue: NS SFQ score: NS **Quality of life outcomes:** VAS degree of well-being: NS **Function outcomes:** SF-36: NS
Diaz-Mitoma, *et al.*, 2003[74] N=15 Poor	CDC (Holmes, 1988 and Fukuda, 1994) criteria	3 months (12 weeks) of treatment	A. Oral Isoprinosine 1 g TID in weeks 1, 3, 5, 7, 9, and 11 only on Monday-Friday; and 1 g/day in weeks 2, 4, 6, 8, 10, and 12 only on Monday-Friday. B. Placebo	**Fatigue:** KPS (% change from baseline): NS **Other:** Activities of daily living scale; no differences but data not provided
McKenzie, *et al.,* 1998 [68] N=60-70 varied by outcome Fair	CDC (Holmes, 1988) and CDC (Fukuda, 1994) criteria	3 months (12 weeks)	A. Oral hydrocortisone 20-30 mg every morning and 5 mg every evening B. Placebo	**Fatigue outcomes:** POMS fatigue subscale: NS POMS vigor subscale: NS **Quality of life outcomes:** Global Wellness scale: NS **Function outcomes:** Activity Scale: NS

Author, year Study type N Quality	Case definition	Duration/ followup	Interventions	Overall Effect: Treatment compared with placebo
Montoya, *et al.,* 2013[71] N=30 Fair	CDC (Fukuda, 1994) criteria	6 months treatment, 6 months followup	A. Oral valganciclovir 900 mg BID for 21 days, then 900 mg/day for total of 6 months B. IV placebo (1% albumen solution) every 30 days for 6 months (6 infusions)	**Fatigue outcomes:** FSS (change in score, negative indicates better health): -0.06 vs. 0.02; p=0.006 MFI-20: NS **Function outcomes:** Self-reported physical function: NS **Other:** CDC Symptom Inventory: NS
Peterson, *et al.,* 1990[70] N=28 Fair	CDC (Holmes, 1988) criteria	6 months	A. IV IgG (1 g/kg) every 30 days for 6 months (6 infusions) B. Placebo	**Functioning:** MOS-SF score for social function higher in placebo group: 5.2 (5.5) vs. 9.4 (7.9); p<0.05 MOS-SF physical: NS
Strayer, *et al.,* 1994[72] N=76-84 varies by outcome Fair	CDC (Holmes,1988) and (Fukuda, 1994) criteria	6 months	A. IV rintatolimod 200 mg twice weekly 4 times, then 400 mg twice weekly for a total of 24 weeks B. Placebo	**Fatigue:** Exercise duration (% change from baseline): 10.3 vs. 2.1; p=0.007 Exercise work (% change from baseline): 11.8 vs. 5.8; p=0.011 **Function:** ADL score (% change from baseline): 23.1 vs. 14.1; p=0.034 KPS score (% change from baseline): +20 vs. 0; p=0.023 **Other:** Decreased used of medications for relief of CFS symptoms declined for rintatolimod but not placebo
Strayer, *et al.,* 2012[73] N=240 Fair to good	CDC (Holmes,1988) and (Fukuda, 1994) criteria	10 months (40 weeks)	A. IV rintatolimod 400 mg twice weekly for 40 weeks B. Placebo	**Fatigue outcomes:** Cardiopulmonary exercise tolerance (% change from baseline: 36.5% vs. 15.2%; p=0.047 **Function outcomes:** KPS score, ADLs, Vitality Score (SF-36), and General Health Perception (SF-36) measured pre and post, but not compared between rintatolimod and placebo groups **Other outcomes:** Decreased use of medications for relief of CFS symptoms: 68% vs. 55%; p=0.048
Wearden, et.al 1998[98] N=69 Fair	Oxford (Sharpe, 1991) criteria	6.5 months	A. Fluoxetine 20 mg/day B. Placebo	**Fatigue:** Chalder Fatigue Scale (mean change from baseline): NS Chalder Fatigue Scale (non-cases of fatigue with score <4): NS **Function:** Functional work capacity (mean change): NS

ADL= Activities of Daily Living; BID=twice a day; CDC= Centers for Disease Control and Prevention; CFS= chronic fatigue syndrome; CGI= Clinical Global Impression change score; FIQ= Fibromyalgia Impact Questionnaire; FSS= Fatigue Severity Scale; g= gram; IgG= immunoglobulin G; IV= intravenous; kg= kilogram; KPS= Karnofsky Performance Scale; MFI-20= Multidimensional Fatigue Inventory; mg= milligram; MOS-SF= Medical Outcome Study Short Form; N= sample size; NS= not significant; POMS= Profile of Mood States; SF-36= 36-item Short Form Survey; SFQ= abbreviated fatigue questionnaire; TID= three times a day; µg= microgram; VAS=visual analogue scale; vs.= versus.

Counseling and Behavior Therapies

Sixteen trials (23 publications) comparing one counseling or behavioral therapy with usual care, wait list control, no treatment, relaxation techniques only, adaptive pacing, anaerobic therapy, GET, or an alternate form of counseling or behavioral therapy met inclusion criteria.[76-98] Twelve trials (17 publications) included only counseling and behavior therapies as the active intervention compared with a control group[76,78-80,83,85,86,88-97] and two trials (4 publications) included an exercise group as a comparison group, in additional to a control group.[81,82,84,98] The results that pertain to the comparisons between the counseling or behavior therapy and "control" groups (defined as wait list, no treatment, usual care, support, relaxation, and adaptive pacing) on outcomes will be discussed in this section (**Table 3** below; **Appendix G4)**, results based on comparisons with exercise programs or GET will be discussed in the Exercise Program section below. One trial included GET with the CBT group compared with a control group[87] and one trial compared face-to-face CBT with telephone CBT and did not include a control group,[77] these studies are discussed in the Head-to-Head and Combination Therapies Sections below. All trials were designed to target symptoms of ME/CFS, not treat the underlying cause. Five trials were rated good-,[90,91,93,95,98] six fair-,[76-78,84,85,88,89] and three poor-quality[80,83,86] (**Appendix H2**).[76,88-90] Half of the trials (50%, 7/14 trials) were of small sample size (n<100) and the duration of illness ranged from 6 months to 52 years. Most trials (57%, 8/14) used the CDC (Fukuda, 1994) criteria to identify people with ME/CFS, while others used the Oxford (Sharpe, 1991) criteria, one developed by Schluederberg in 1992, and one study instead used a combination of a CFS questionnaire, psychiatric assessment, and medical assessment to rule out other conditions and diagnose ME/CFS.[84]

Adherence was not reported in nine of the trials,[76,78,80,83,85,86,89,93,98] one trial only reported that the adherence was "good,"[91] while another trial stated that all completed CBT.[90] One trial reported that the participants completed an average of 10 out of 13 counseling sessions, but no other information,[84] and two trials reported contamination, not specifically adherence (6% in CBT group received support, while 8% in support group received CBT[88] and 3% in counseling group received support, while 1% in support group received pragmatic rehabilitation and 10% in support did not receive any treatment).[95] Attrition was reported in most trials and was generally low (<20%) and similar between groups. Only three trials reported high or differential attrition, and the study with the highest attrition in all groups (41% vs. 35% vs. 23%) had 14 months of followup,[89] while most others were no longer than 12 months. One trial conducted a stepped care approach, providing participants with a self-instruction program to follow for 16 weeks or delayed care followed by CBT, and found that after 16 weeks, 57 percent of those in the self-instruction group and 22 percent in the delayed group did not want to continue on to CBT, and therefore dropped from the study.[92] Major limitations of trials include enrolling fewer than 20 subjects in an arm,[83] high loss to followup,[77] unclear if use of intention-to-treat analysis,[80,83,84,86] unclear randomization or allocation process,[76,80,83,84,86,91] more men allocated to the intervention group,[88] and lack of or unclear information about blinding of outcome assessor.[76,77,80,84,85,90,91,93] Due to the nature of the interventions, most trials could not blind patients or care providers. Trials were either funded by government or organizational grants (fully or in part)[76,80,83,84,86,88,90,91,93,95,98] or the funding source was not reported.[77,85]

There is low strength evidence, based on 14 trials, that CBT, either group or individual; self-instruction booklets; pragmatic rehabilitation; peer-to-peer counseling; and symptom consultation provide improvement in fatigue, function, quality of life, and employment in adult patients with ME/CFS. When combining all studies comparing any type of counseling to no

treatment, support, relaxation, or adaptive pacing there is low strength of evidence that counseling improves fatigue (8/12 trials showed positive effect), functioning measures (5/12 trials showed positive effect; 2/12 showed mixed results on measures), quality of life (2/5 trials showed positive effect; 1/5 showed mixed results on measures), and global improvement (3/3 trials showed positive effect). Harms of counseling and behavioral therapies were poorly reported but there is low strength of evidence that CBT is not associated with harms based on one moderate-sized trial.

Twelve trials (n=1,889) of counseling compared with no treatment, support, relaxation, or adaptive pacing, reported overall functioning measured by the SF-36 physical functioning subscale, KPS, SIP-8, and the functional impairment scale.[76,78,80,83-85,88-90,92,95,98] Results were mainly positive, but mixed. In five trials[78,84,89,90,98] counseling improved overall functioning compared with controls on various measures, while two trials reported mixed results using different measures in the same study,[85,88] one trial reported improvement in the control group compared with counseling,[76] and the other four trials reported no differences between groups.[80,83,92,95]

Eight trials used the SF-36 physical functioning subscale to measure overall functioning[78,83-85,88,92,95,98] and results were mixed. Four trials[78,84,85,98] reported significantly more improvement in the CBT group compared with controls (71% vs. 49% improved by ≥8 points from baseline at 1 year, p=0.0068[98] and 63% vs. 17% with score >83 at 6 months followup, p<0.001[78]) or better scores (mean scores of 58.64 vs. 39.72 at 1 year, p<0.01[84] and mean scores of 65.9 vs. 60.2 at about 6-12 months, p=0.011[85]). However, by 5 years in one trial the results were no longer significantly different,[79] and in another trial on the SIP-8 the outcome was reversed with worse functioning reported in the self-instruction group compared with the wait list control (mean scores of 1,515 vs. 1,319, p<0.001).[85] The other four trials reported no differences between the counseling group and controls.[83,88,92,95] However, one trial also measured functioning using a walking speed test and found improved walking speed in the CBT group compared with controls (difference from baseline to 12 months for CBT vs. support: 1.77; 95% CI, 0.025 to 3.51; p=0.0055 and difference from baseline to 12 months for CBT vs. no intervention: 2.83; 95% CI, 1.12 to 5.53; p=0.0055).[88] Another trial conducted a post hoc analysis on those with baseline functional disability (defined as SF-36 physical functioning baseline score ≤70) and reported a significant improvement in the self-instruction group compared with the wait list control (mean change from baseline CBT vs. control: 9.05; 95% CI, 0.2 to 17.9; p<0.05).[92] When trials using the SF-36 physical functioning subscale were pooled there was a significant effect for the intervention group to have better scores compared with controls at followup; weighted mean difference of 7.73 (95% CI, 3.58 to 11.87, **Figure 3**). Even when one outlier that showed a more significant difference than the other trials[78,79] was removed the difference was still significant with a weighted mean difference of 7.18 (95% CI, 4.53 to 9.83).

Two trials used the KPS to measure overall functioning[89,90] and both reported significantly more improvement in the CBT group compared with controls (73% vs. 23% improved by ≥10 point at 12 months; difference of 50% CBT vs. no intervention; 95% CI, 28 to 72%[90] and 49% vs. 19% in support vs. 23% in no intervention improved by ≥10 point as well as had a final score of ≥80 at 14 months; p=0.001[89]). One trial also reported a significant improvement in the CBT group compared with controls on the SIP-8 at 14 months (treatment effects: 263; 95% CI, 38 to 488; p=0.0223 for CBT vs. support and 222; 95% CI, 3 to 441; p=0.0470 for CBT vs. control.[89]

One trial only used the SIP-8 to measure overall functioning[76] and reported worse functional impairment in the CBT group compared with the wait list control (mean change from baseline at

6 months: 29 vs. -293; p=0.004). Along with the other outcomes reported above the SIP-8 showed mixed results. One trial only used the functional impairment scale to measure overall functioning and found no differences between counseling and the wait list control at 6 months followup.[80]

Twelve trials (n=1,887) of counseling compared with no treatment, support, relaxation, or adaptive pacing, reported decreased fatigue measured by the Chalder Fatigue Scale, FSS, CIS, POMS-fatigue, Profile of fatigue-related symptoms scale, and the SF-36 vitality subscale.[76,78,80,83-86,88,89,92,95,98] Results were primarily positive, but mixed. In nine trials counseling significantly decreased fatigue compared with controls on various measures,[78,80,83-85,88,89,92,93,98] while the other three trials showed no differences between groups.[76,86,95]

Four trials used the Chalder Fatigue Scale to measure fatigue[78,88,95,98] and results were primarily positive. Three trials reported significantly more decreases in fatigue in the counseling group compared with the controls (63% vs. 15% were non-cases of fatigue with a score <4 at 6 months; p=0.001[78]) or better scores (difference in scores from baseline at 1 year for CBT vs. support: -3.16; 95% CI, -5.59 to -0.74; p=0.011 and CBT vs. no intervention: -2.61; 95% CI, -4.92 to -0.30; p=0.027[88] and at 52 weeks for CBT vs. no intervention: -3.4; 95% CI, -5.0 to -1.8; p=0.0001[98]). However, by 5 years in one trial the results were no longer significantly different.[79] The other trial that used the Chalder Fatigue Scale reported statistically significantly better fatigues scores in the pragmatic rehabilitation group than the usual care group (treatment effect estimate of -1.18; 95% CI, -2.18 to -0.18; p=0.021), but by 70 weeks there were no differences.[95] This study conducted post hoc analyses to determine what may predict change on the Chalder Fatigue Scale and found significant effect for age (-0.10; 95% CI, -0.19 to -0.003; p=0.044), duration of illness (-0.01; 95% CI, -0.02 to -0.003; p=0.008), and those with severe problems as measured by the EQ-5D mobility scale (-2.95;95% CI, -5.51 to -0.40; p=0.024). Meaning those who were younger, had shorter illness durations, and less severe mobility problems at baseline showed greater improvements in fatigue at 70 weeks.[96] Due to the variability in how the Chalder fatigue scale was used across studies, these results could not be pooled.

Four trials used the CIS to measure fatigue[76,85,89,92] and results were primarily positive. Three trials reported significantly more decreases in fatigue in the counseling groups compared with the controls (35% vs. 13% support vs. 17% no intervention improved at 14 months; p=0.009 for CBT vs. support, p=0.026 for CBT vs. control[89], 27% vs. 7% improved at about 6-12 months; OR 4.9; 95% CI, 1.9 to 12.9; p<0.001,[85] and 33% vs. 9% at 6 months; OR 5.0; 95% CI< 1.69 to 14.57;[93] all used a reliable change score of >1.64 and final score of ≤36 to indicated improved). Only one non randomized trial found no differences between groups at 6 months.[76]

Two trials used the FSS to measure fatigue[83,84] and both found significantly lower fatigue scores in the counseling groups compared with the controls (mean scores of 52.9 vs. 59.4 at 4 months; p=0.04[83] and mean scores of 5.37 vs. 5.62 at 1 year, but p value not reported[84]). One trial conducted post hoc analyses based on whether or not individuals stayed within their energy envelope, meaning they avoided overexertion and under exertion by exerting a comfortable range of energy, or strayed outside their energy envelope.[82] Individuals rated their perceived energy and expended energy and this was used to determine which individuals stayed within their energy envelope (n=49) and which were outside their energy envelope (n=32). At 12 months there was a statistically significant improvement in mean fatigue and functioning scores from baseline for those who stayed within their energy envelope compared with those who were outside their energy envelope (fatigue scores: -0.9 vs. 0.1; p<0.01 and functioning scores: 17 vs. 0; p=0.03). The second additional analysis compared fatigue and functioning outcomes based on

homework compliance.[81] They identified three groups based on the amount of homework completed; minimum compliance completed 0 to 25 percent, moderate compliance completed 25.1 to 75 percent, and maximum compliance completed 75.1 to 100 percent of their assigned homework. When they assigned individuals to groups they noted that the highest percentage in the maximum group (56%) were in the cognitive therapy group, the highest percentage in the moderate group (34%) were in the CBT group, and the highest percentage in the minimum group (38%) were in the anaerobic and relaxation groups. At 12 months, though there was a trend toward better improvement in fatigue and functioning scores for the maximum compliance group compared with the other groups this did not reach significance.

One trial used only the Profile of fatigue-related symptoms scale to measure fatigue[80] and reported better scores in the counseling group compared with the wait list control (mean scores of 2.68 vs. 3.84 at 6 months; p=00.04), while one trial reported no differences in fatigue scores using the POMS.[86]

Five trials (n=568) of counseling compared with no treatment, support, or relaxation, reported quality of life measured by the QOLI, QLI, QLS, EuroQol, and the health utilities index.[84,86,88,89,91] Results were mixed, but primarily positive. Two trials reported better scores in the counseling group compared with controls (mean QOLI scores at 12 weeks: 2.81 vs. 3.26; p=0.02[86] and mean change in QLI scores from baseline at 12 months: 2.6 vs. 0.6; p<0.05[91]), one trial[89] reported better quality of life compared with support but not no treatment on the EuroQol at 14 months (treatment effects for CBT vs. support: -9.2; 95% CI, -15.6 to -2.8; p=0.0049 and treatment effects for CBT vs. no treatment: -2.3; 95% CI, -8.4 to 3.8; p=0.4619), one trial[84] reported slightly better scores on the QLI in the cognitive group compared with CBT and controls at 1 year, but the p value was not reported (69.10 for CBT vs. 72.52 for cognitive vs. 63.00 for anaerobic activity vs. 72.00 for relaxation), and the final study reported no differences between groups.[88]

Five trials (n=1,065) of counseling compared with no treatment, support, relaxation, or adaptive pacing reported employment outcomes including proportion working full- or part-time, hours worked per week or per 24-hour period, and level of work impairment measured by the Work and Social Adjustment Scale.[76,78,79,84,89,98] Results were primarily positive. Both trials measuring work impairment with the Work and Social Adjustment Scale reported significantly better scores for the CBT group compared with controls (mean scores of 3.3 vs. 5.4 at 6 months; p<0.001 on scale scored with range 0-8;[78] mean scores of 21.0 vs. 24.5 at 1 year; p=0.0001 on scale scored with range 0-45[98]). Three trials reported the number of hours, either per week or per 24-hours, individuals were working, with one trial reporting significantly more hours worked per week for the CBT group compared with relaxation (mean hours of 35.57 vs. 24.00 at 5 years; p<0.04),[79] however, one trial[89] reported significantly more hours worked on 24-hour timetable in CBT group compared with support group but not the no treatment group (treatment effects for CBT vs. support: -9.6; 95% CI, -17.1 to -2.0; p=0.0132 and CBT vs. no treatment: -5.9; 95% CI, -13.2 to 1.4; p=0.1134), and one trial reported no differences between groups.[76] Two trials reported no differences in the proportion of individuals working full- or part-time at 1 year[84] or 5 years.[79]

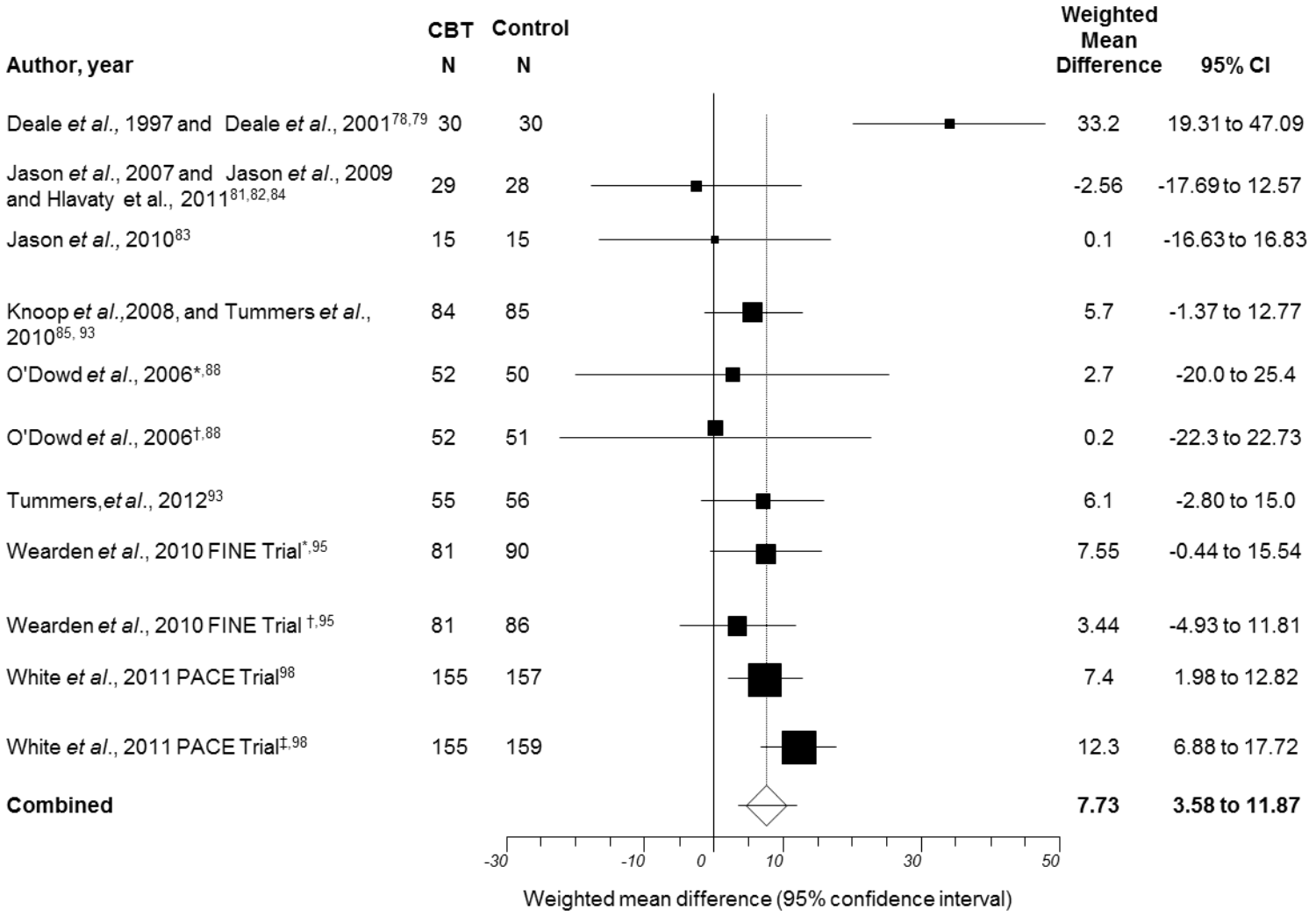
Three trials (n=886) of counseling compared with no treatment, relaxation, or adaptive pacing reported global improvement using the CGI, spontaneous reporting of fully recovered or feeling much better, relapses, full recovery, and no longer meeting ME/CFS criteria.[78,79,89,98] All three trials reported better global improvement for counseling compared with controls. Significantly more individuals reported being fully recovered or that they felt much better in the

CBT group compared with the support group and no intervention group (50% vs. 15% vs. 32%; p<0.001 for CBT vs. support, p=0.034 for CBT vs. no intervention).[89] Significantly more individuals in the CBT group reported improvement compared with controls in two trials (70% vs. 31%; p<0.01[78] and 41% vs. 31%; p=0.013[98]). One trial[79] also followed up 5 years after counseling and continued to report more improvement in the CBT group compared with relaxation (68% vs. 43% with symptoms "steadily improved" not "consistently absent' or "mild"; p=0.05; 24% vs. 4% with complete recovery; p=0.04; 36% vs. 7% with no relapses; p=0.02; and mean number of relapses of 2.58 vs. 4.08; p<0.01), however there was no difference in the number of individuals currently meeting the Oxford (Sharpe, 1991) criteria for ME/CFS (52% vs. 39%; p=0.42).

Only three good-quality trials reported anything about harms after counseling or behavior therapies. One large (n=630) trial reported fewer total harms in the CBT group (848) compared with adaptive pacing (949) and no treatment (977), but the p value was not reported, the same trial also reported fewer serious harms in the CBT group per 100 person-years (5.0; 95% CI, 2.2 to 9.8) compared with adaptive pacing (10.1; 95% CI, 5.8 to 16.3), but was similar to no treatment (4.4; 95% CI,1.8 to 9.0).[98] One trial reported none withdrew due to harms[91] and the other trial reported no differences between groups for reported harms or withdrawals due to harms,[95] but no other information was provide about harms.

In summary most trials of CBT or other counseling techniques suggested improvement in overall functioning and fatigue symptoms in ME/CFS patients though in a trial that followed individuals up 5 years after counseling, this affect was no longer seen. In addition, studies used various measures to detect decreases in fatigue or improvements in overall functioning and results were often similar based on the measure being used, but difficult to compare across measures. Also, few studies reported the clinical significance, if available, of the improvement in scores. Harms were rarely reported.

Figure 3. Meta-analysis of mean changes in SF-36 physical function subscale scores for CBT compared with



controls

*Using support as the comparison

†Using usual care as the comparison

‡ Using adaptive pacing as the comparison

CBT= cognitive behavioral therapy CI= confidence interval; N= sample size; SF-36= Short Form 36-item survey.

Table 3. Trials of counseling and behavioral therapies for ME/CFS

Author, year Study type N Quality	Case definition	Duration/ followup	Interventions	Overall effect Intervention A vs. intervention B vs. intervention C etc.
Bazelmans, *et al.,* 2005[76] Non-randomized trial N=65 Fair	CDC (Fukuda, 1994)	6 months	A: Group CBT B: Wait list control	**Fatigue outcomes:** CIS: NS **Function outcomes:** Functional impairment improved in control group on SIP-8 at 6 months, mean change in scores from baseline: 29 vs. -293; p=0.004 **Employment:** Hours worked/week: NS

Author, year Study type N Quality	Case definition	Duration/ followup	Interventions	Overall effect Intervention A vs. intervention B vs. intervention C etc.
Deale, *et al.,* 1997[78] N=60 Deale*, et al.,* 2001[79] N=53 Fair	Oxford (Sharpe, 1991) and United States (Schluederberg, 1992) criteria	Deale, 1997: 6 months Deale, 2001: 5 years	A. CBT B. Relaxation	**Fatigue outcomes:** Fatigue rating by assessor at 3 months as "better or much better" higher in the CBT group: 72% (18/25) vs. 17% (4/23); p<0.001 Non-cases of fatigue (score <4 on Chalder Fatigue Scale) at 6 months higher in the CBT group: 63% (17/27) vs. 15% (4/26); p=0.001 5 year followup: NS **Function outcomes:** % With good outcome on SF-36 physical functioning subscale better in CBT group at 6 months: 63% (19/30) vs. 17% (5/30); difference of 46% (95% CI, 24 to 68), p<0.001 5 year followup: NS Functioning rating by assessor at 3 months as "better or much better" higher in CBT group: 80% (20/25) vs. 26% (6/23); p<0.001 **Employment outcomes:** Work and social adjustment scale subscale scores better in CBT at 6 months, mean (SD): 3.3 (2.2) vs. 5.4 (1.8) p<0.001, between group differences over time Hours worked per week at 5 years was higher in CBT group, mean (SD): 35.57 (8.11) vs. 24.00 (4.97); p<0.04 % With full- or part-time employment at 5 year followup: NS **Other outcomes:** Global improvement rating "better or much better" higher in the CBT group at 6 months: 70% (19/27) vs. 31% (8/26); p<0.01 Global improvement rating "better or much better" higher in the CBT group at 5 years: 68% (17/25) vs. 36% (10/28); p=0.05 **Outcomes at 5 year followup:** Symptoms "steadily improved" not "consistently absent' or "mild" higher in the CBT group: 68% (17/25) vs. 43% (12/28); p=0.05 Complete recovery higher in the CBT group: 24% (6/25) vs. 4% (1/28); p=0.04 No relapses higher in the CBT group: 36% (9/25) vs. 7% (2/28); p=0.02 Fewer number of relapses in CBT group, mean (SD): 2.58 (2.21) vs. 4.08 (1.55); p<0.01 No longer meeting U.K. criteria for CFS: 52% (13/25) vs. 39% (11/28); p=NS
Goudsmit, *et al.,* 2009[80] N=44 Poor	Oxford (Sharpe, 1991) criteria	6 months	A. Counseling B. Wait list	**Fatigue outcomes:** Profile of fatigue-related symptoms scale scores better in counseling group at 6 months, mean (SD): 2.68 (1.41) vs. 3.84 (1.40); p=0.04 **Function outcomes:** Functional impairment scale scores: NS

Author, year Study type N Quality	Case definition	Duration/ followup	Interventions	Overall effect Intervention A vs. intervention B vs. intervention C etc.
Jason, *et al.,* 2007[84] Jason, *et al.,* 2009[82] Hlavaty, *et al.,* 2011[81] N=114 Fair *Same study on Table 5 and 6*	CFS Questionnaire, psychiatric assessment for DSM-IV diagnosis, and medical assessment	12 months	A: CBT B: COG C: ACT D. Relaxation	**Fatigue outcomes:** Fatigue scores better in CBT group for FSS scores at 12 months, mean (SD): 5.37 (1.19) vs. 5.87 (1.01) vs. 5.77 (1.43) vs. 5.62 (1.06); p=NR Comparison by energy envelope for those who stayed within envelope vs. outside envelope at 12 months was 5.3 vs. 6.3. Change at 12 months from baseline: -0.9 vs. 0.1; p<0.01 The comparison by homework compliance level, change in score at 12 months from baseline: -0.17 (0.73) vs. -0.51 (1.00) vs. -0.54 (1.09); p=NR **Quality of life outcomes:** Quality of life slightly better in COG group based on QLS scores at 12 months mean (SD): 69.10 (18.99) vs. 72.52 (10.84) vs. 63.00 (13.86) vs. 72.00 (19.70); p=NR **Function outcomes:** Functional scores better in CBT, COG, and relaxation group than ACT on SF-36 physical functioning subscale scores at 12 months, mean (SD): 58.64 (30.44) vs. 61.09 (23.74) vs. 39.72 (27.63) vs. 61.20 (27.70) p<0.01, for CBT and COG over time vs. ACT over time Comparison by energy envelope for those who stayed within envelope vs. outside envelope at 12 months: 65 vs. 43, change at 12 months from baseline: 17 vs. 0; p=0.03 Comparison by homework compliance level, change in score at 12 months from baseline: 6.99 (19.30) vs. 7.55 (18.85) vs. 17.50 (18.09); p=NR % Achieving clinically significant improvement: NS **Employment outcomes:** % Employed at 12 month followup: NS
Jason, *et al.,* 2010[83] N=30 Poor	CDC (Fukuda, 1994) criteria	4 months	A. Buddy counseling B. Control, no treatment for 4 months	**Fatigue outcomes:** FSS scores better in buddy counseling group at 4 months, mean (SD): 52.9 (10.5) vs. 59.4 (3.7); p=0.04 SF-36 vitality subscale scores better in buddy counseling group at 4 months, mean (SD): 29.3 (13.9) vs. 24.7 (9.7); p<0.05 **Function outcomes:** SF-36 physical functioning subscale: NS

Author, year Study type N Quality	Case definition	Duration/ followup	Interventions	Overall effect Intervention A vs. intervention B vs. intervention C etc.
Knoop, *et al.,* 2008[85] Tummers, *et al.,* 2010[92] N=169 Fair	CDC (Fukuda, 1994) criteria	6-12 months depending on length of treatment	A. Self-instruction B. Wait list control Tummers, 2010 A. Stepped care B. Usual care	**Fatigue outcomes:** CIS fatigue severity scores better in self-instruction group at second assessment, mean (SD): 38.9 (12.1) vs. 46.4 (8.7); p<0.001 % With reduction in CIS fatigue severity scores higher in self-instruction group: 27% (23/84; 95% CI 18 to 37) vs. 7% (6/85; 95% CI 2 to 13); OR 4.9 (95% CI, 1.9 to 12.9); p<0.001 **Function outcomes:** SF-36 physical functioning subscale better in self-instruction group at second assessment, mean (SD): 65.9 (23.2) vs. 60.2 (23.7); p=0.011 Functional impairment SIP-8 scores worse in self-instruction group at second assessment, mean (SD): 1,515 (545) vs. 1,319 (619); p<0.001 *Tummers, 2010 additional stepped care vs. usual care analysis* **Fatigue outcomes:** CIS fatigue severity scores: NR % With reduction in CIS fatigue severity scores: NR **Function outcomes:** SF-36 physical functioning subscale: NS **Other outcomes:** Number of CBT sessions for stepped care vs. usual: 10.9 (4.4) vs. 14.5 (5.3); p<0.01 Median minutes in sessions (range): 420 (120-1,440) vs. 720 (120-2,040); p=0.01
Lopez, *et al.,* 2011[86] N=58 Poor	CDC (Fukuda, 1994) criteria	3 months (12 weeks)	A. Group CBT B. Control, 1 session of psychoeducation summarizing strategies	**Fatigue outcomes:** POMS-Fatigue subscale: NS **Quality of life outcomes:** Category scores with lower scores indicating better health after treatment better in group CBT, mean (SD): 2.81 (1.15) vs. 3.26 (0.87); p=0.02 Raw score after treatment: 1.17 (1.83) vs. 0.82 (1.37); p=0.05 T score after treatment: 39.28 (14.17) vs. 36.42 (10.56); p=0.05

Author, year Study type N Quality	Case definition	Duration/ followup	Interventions	Overall effect Intervention A vs. intervention B vs. intervention C etc.
O'Dowd, *et al.,* 2006[88] N=153 Fair	CDC (Fukuda, 1994) criteria	12 months	A. Group CBT B. Group support C. Usual care	**Fatigue outcomes:** Fatigue difference between groups from baseline for CBT vs. support at 12 months: -3.16 (95% CI, -5.59 to -0.74); p=0.011 Fatigue difference between groups from baseline for CBT vs. usual care at 12 months: -2.61 (95% CI, -4.92 to -0.30); p=0.027 **Function outcomes:** Normal walking speed higher in CBT group on mean incremental shuttle walking test; at 6 and/or 12 months: 11.58 (0.71) vs. 9.82 (0.53) vs. 8.76 (0.47); p=0.006 Difference between groups from baseline to 12 months for CBT vs. support: 1.77 (95% CI, 0.025 to 3.51); p=0.0055 Difference between groups from baseline to 12 months for CBT vs. usual care: 2.83 (95% CI, 1.12 to 5.53); p=0.0055 SF-36 physical functioning subscale: NS **Quality of life outcomes:** Health related quality of life utility scores: NS

Author, year Study type N Quality	Case definition	Duration/ followup	Interventions	Overall effect Intervention A vs. intervention B vs. intervention C etc.
Prins, *et al.,* 2001[89] N=196 Fair	CDC (Fukuda, 1994) criteria, except for the requirement of 4/8 additional symptoms to be present	14 months	A. CBT B. Support group meetings C. No intervention	**Fatigue outcomes:** % With improvement on CIS scores higher in CBT group at 14 months: 35% (20/58) vs. 13% (8/62) vs. 17% (13/76); p=0.009 for CBT vs. support, p=0.026 for CBT vs. control Treatment effects on CIS scores for CBT vs. support at 14 months: 5.8 (95% CI, 2.2 to 9.4); p=0.0015 Treatment effects on CIS scores for CBT vs. control at 14 months: 5.6 (95% CI, 2.1 to 9.0); p=0.0016 **Function outcomes:** % With improvement on KPS score better in CBT group at 14 months: 49% (28/57) vs. 19% (12/62) vs. 23% (17/75); p=0.001 Treatment effects for CBT vs. support on KPS score at 14 months: -6.3 (95% CI, -9.6 to -3.0); p=0.0002 Treatment effects for CBT vs. control on KPS at 14 months: -5.4 (95% CI, -8.6 to -2.2); p=0.0009 Treatment effects for CBT vs. support on SIP-8 score at 14 months: 263 (95% CI, 38 to 488); p=0.0223 Treatment effects for CBT vs. control on SIP-8 at 14 months: 222 (95% CI, 3 to 441); p=0.0470 **Quality of life outcomes:** Fatigue treatment effects for CBT vs. support on EuroQol scale at 14 months: -9.2 (95% CI, -15.6 to -2.8); p=0.0049 Treatment effects CBT vs. control on EuroQol scale: NS **Employment outcomes:** Treatment effects for CBT vs. support on hours worked on 24-hour timetable at 14 months: -9.6 (95% CI, -17.1 to -2.0); p=0.0132 Treatment effects for CBT vs. control on hours worked on 24-hour timetable at 14 months: NS **Other outcomes:** % Of patient reporting they were fully recovered or felt much better based better in CBT group at 14 months: 50% (29/58) vs. 15% (9/62) vs. 32% (24/76); p<0.001 for CBT vs. support, p=0.034 for CBT vs. control
Sharpe, *et al.,* 1996[90] N= 60 Good	Oxford (Sharpe 1991) criteria	12 months	A. CBT B. Usual care	**Function outcomes:** Functional scores of ≥80 on KPS better in CBT group at 12 months: 73% (22/30) vs. 27% (8/30); difference of 47% (95% CI, 24 to 69) Improvement of ≥10 points on KPS better in CBT group at 12 months: 73% (22/30) vs. 23% (7/30); difference of 50% (95% CI 28 to 72%)

Author, year Study type N Quality	Case definition	Duration/ followup	Interventions	Overall effect Intervention A vs. intervention B vs. intervention C etc.
Taylor, 2004[91] N=47 Good	CDC (Fukuda, 1994) criteria	12 months	A. Counseling B. Wait list	**Quality of life outcomes**: QLI scores better in counseling group at 12 months, mean (SD): 15.7 (3.7) vs. 14.6 (4.1); mean change from baseline: 2.6 vs. 0.6; p<0.05 Health and function subscale at 12 months: 14.1 (1.7) vs. 13.6 (1.8) Social and economic subscale at 12 months: 15.6 (0.8) vs. 15.5 (0.9) Psychological and spiritual subscale at 12 months: 15.5 (1.1) vs. 15.1 (1.2) Family subscale at 12 months: 15.6 (0.8) vs. 15.5 (0.9); mean change from baseline: 0.2 vs. -0.2; p<0.05
Tummers, *et al.,* 2012[93] N=111 Good	CDC (Fukuda, 1994) criteria	6 months	A. Self-instruction B. Wait list	**Fatigue outcomes:** Fatigue severity scores better in self-instruction group on CIS fatigue scale at second assessment, mean (SD): 39.6 (14.1) vs. 48.3 (8.1); p<0.01 % With reduction in CIS fatigue severity scores: 33% (18/55) vs. 9% (5/56); OR 5.0 (95% CI, 1.69 to 14.57) Of those within the disabled range at baseline (SF-36 physical functioning subscale score ≤70, n=99), those in the self-instruction group improved more at the second assessment, mean change from baseline: -12.4 vs. -2.4; difference: -9.9 (95% CI, -5.4 to -14.3); p<0.01 **Function outcomes:** SF-36 physical functioning subscale (main analysis): NS Of those within the disabled range at baseline (SF-36 physical functioning subscale score ≤70, n=99), those in the self-instruction group improved more at the second assessment, mean change from baseline: 18.5 vs. 9.6, difference: 9.05 (95% CI, 0.2 to 17.9); p<0.05
Tummers, *et al.,* 2013[94] "See Knoop, 2008 and Tummers, 2012" (Fair) Secondary analysis of Knoop, et al., 2008 & Tummers, et al., 2012 combined	CDC (Fukuda, 1994) criteria	6-12 months (based on the RCTs)	A. Self-instruction B. Wait list	**Fatigue outcomes:** *Interaction tests for potential moderators from linear regression models (95% CI):* Age (years): 0.15 (0.01 to 0.045); p<0.05 Depression: 0.15 (0.04 to 1.95); p=0.04 Avoidance of activity: 0.17 (0.03 to 1.78); p=0.04 Perpetuating factors: self-efficacy: NS, somatic attribution: NS, focus on bodily symptoms: NS *Interaction tests for potential moderators from logistic regression models (95% CI):* Avoidance of activity: 1.34 (1.03 to 1.74); p=0.03 Depression: 1.40 (1.08 to 1.82); p=0.01 Age (years): NS Perpetuating factors: self-efficacy: NS, somatic attribution: NS, focus on bodily symptoms: NS

Author, year Study type N Quality	Case definition	Duration/ followup	Interventions	Overall effect Intervention A vs. intervention B vs. intervention C etc.
Wearden, *et al.,* 2010[95] ***FINE Trial*** Wearden, *et al.,* 2012[96] Wearden, *et al.,* 2013[97] N=257 Good	Oxford (Sharpe,1991) criteria	4.5 months (18 weeks) treatment; 17.5 months (70 weeks) total followup	A. Pragmatic rehab B. Supportive listening C. Usual care	**Fatigue outcomes:** Chalder Fatigue Scale scores at 70 weeks: NS *Significant regression coefficients for interaction between putative moderators and treatment:* HADS baseline depression score: -0.67 (95%, CI -1.25 to -0.10); p=0.022 HADS baseline total score: -0.30 (95% CI, -0.58 to -0.02); p=0.039 EQ-5D self-care scale, those with severe problems: -28.72 (95% CI, -32.14 to -25.31); p<0.001 *Significant regression coefficients to predict change in Chalder Fatigue Scale scores:* Age: -0.10 (95% CI, -0.19 to -0.003); p=0.044 Duration of illness: -0.01 (95% CI, -0.02 to -0.003); p=0.008 EQ-5D mobility scale; those with severe problems: -2.95 (95% CI, -5.51 to -0.40); p=0.024 **Function outcomes:** Functional scores better in usual group on SF-36 physical functioning subscale at 20 weeks, mean (SD): 39.94 (25.21) vs. 33.28 (22.94) vs. 40.27 (26.45); treatment effect estimate -7.54; 95% CI, -2.96 to -0.11; p=0.035 for supportive listening vs. usual care At 70 weeks: NS

Author, year Study type N Quality	Case definition	Duration/ followup	Interventions	Overall effect Intervention A vs. intervention B vs. intervention C etc.
White, *et al.,* 2011[98] ***PACE Trial*** N=480 Good *Same study on Table 5 and 6*	Oxford (Sharpe, 1991) criteria	13 months (52 weeks)	A. CBT B. APT D. Usual care	**Fatigue outcomes:** Fatigue scores better in CBT group on Chalder Fatigue Scale at 52 weeks, mean (SD): 20.3 (8.0) vs. 23.1 (7.3) vs. 23.8 (6.6) Mean difference CBT vs. APT at 52 weeks (95% CI): -2.7 (-4.4 to -1.1) p=0.0027 Mean difference CBT vs. usual care at 52 weeks (95% CI): -3.4 (-5.0 to -1.8) p=0.0001 % Improved from baseline (by ≥2 points) was higher in the CBT group: 76% (113/148) vs. 65% (99/153) vs. 65% (98/152) **Function outcomes:** Functional scores better in CBT group on SF-36 physical functioning subscale at 52 weeks, mean (SD): 58.2 (24.1) vs. 45.9 (24.9) vs. 50.8 (24.7) Mean difference CBT vs. APT at 52 weeks (95% CI): 10.5 (5.4 to 15.6) p=0.0002 Mean difference CBT vs. usual care at 52 weeks (95% CI) : 7.1 (2.0 to 12.1) p=0.0068 % Improved from baseline (by ≥8 points) was higher in CBT group: 71% (105/148) vs. 49% (75/153) vs. 58% (88/152) **Employment outcomes:** Work and social adjustment scale scores better in CBT group at 52 weeks, mean (SD): 21.0 (9.6) vs. 24.5 (8.8) vs. 23.9 (9.2); p=0.0001 for CBT vs. control and CBT vs. APT **Other outcomes:** More with positive change in CBT group on self-rated CGI at 52 weeks: 41% (61/147) vs. 31% (47/153) vs.25% (38/152) Minimum change: 52% (77/147) vs63% (96/153) vs.66% (100/152) Negative change: 6% (9/147) vs. 7% (10/153) vs. vs. 9% (14/152) *The positive change vs. negative change, OR (95% CI)* CBT vs. APT: 1.7 (1.0 to 2.7) p=0.034 CBT vs. usual care: 2.2 (1.2 to 3.9) p=0.011

ACT= anaerobic activity therapy; APT= adaptive pacing therapy; CBT= cognitive behavioral therapy; CDC= Centers for Disease Control and Prevention; CFS= chronic fatigue syndrome; CI= confidence interval; CIS= Checklist of Individual Strength; COG= cognitive therapy; DSM-IV= Diagnostic and Statistical Manual fourth edition; FSS= fatigue severity scale; HADS= Hospital Anxiety and Depression Scale; KPS= Karnofsky Performance Scale; N= sample size; NR= not reported; NS= not significant; OR= odds ratio; POMS= Profile of Mood States; QLI= Quality of Life Index; QLS= Quality of life scale; RCT= randomized controlled trial; SD= standard deviation; SF-36= 36-item Short Form Survey; SIP-8= Sickness Impact Profile 8-Items; U.K.= United Kingdom; vs. = versus.

Complementary and Alternative Medicine Therapies

Seven trials comparing one CAM approach with usual care, placebo, or alternative CAM intervention (**Table 4** below; **Appendix G4**).[99-105] Three trials aimed at treating a potential underlying pathology (insulin-like growth factor, antioxidant, acetyl-carnitine deficiency)[100-102] whereas the others targeted symptoms. Two trials were of good-quality [101,103], four fair-[99,102,104,105] and one poor-quality[100] (**Appendix H2**). Trials evaluated different dietary approaches or supplements, distant healing, homeopathy, melatonin, and phototherapy. Most were conducted in Europe and all but one[103] of the trials were of small sample size (n<100). Major limitations of studies include inadequate or unclear randomization,[100,105] enrolling fewer than 20 subjects in an arm,[99,100] high loss to followup,[99] lack of intention-to-treat analysis of outcomes,[99,104,105] unclear or inadequate blinding,[99,102,105] and groups not similar at baseline.[101] Trials were either funded by foundations or trusts,[103-105] pharmaceutical companies (fully or in part)[101,102] or the funding source was not reported .[99,100]

The evidence on CAM interventions was insufficient to draw conclusions as interventions included only single trials, were of small sample size, and most had significant methodological limitations. Two fair-quality trials, one with homeopathy and one with L-carnitine preparations, found improvement in some measure of fatigue and/or function, with no differences found in other measures. All other trials of CAM interventions found no significant improvements compared with placebo, usual care, or an alternative CAM approach. Adherence was low in one trial of a low sugar/low yeast diet but otherwise adherence and harms were not well reported.

One good-quality trial (n=57) compared *Acclydine©,* a combination of amino acids and a food supplement derived from the plant *Solanum dulcamara* proposed to increase biologically active insulin-like growth factor (IGF-1), with an identical placebo.[101] Patients were identified based on the CDC (Fukuda, 1994) criteria and followed over 14 weeks for measures of fatigue and function. Although adherence was not reported, they found no differences in fatigue severity based on the CIS fatigue severity subscale questionnaire (1.1; 95% CI, -4.4 to 6.5) or self assessed daily fatigue level (-0.2, 95% CI, -1.2 to 0.9). They also found no difference in function based on the SIP-8 (59.1,95% CI -201.7 to 319.8), and physical activity measured with an actometer motion-sensing device (4.1, 95% CI -5.9 to 14.0).[101] Of note, they also found no difference in IGF-1 blood levels between groups. Attrition was low and no harms were reported.

One fair-quality study (n=89) compared acetyl-L-carnitine (2 g/day) with propionyl-L-carnitine (2 g/day) and with a combination of both.[102] Patients were eligible based on CDC (Fukuda, 1994) criteria and outcomes included CGI and MFI-20 comparing scores 8 weeks prior to intervention and after 24 weeks of treatment. They had 20 percent attrition (18 of 90 enrolled) and did not report adherence but found improvement in CGI for acetyl-l-carnitine (59%) and propionyl-L-carnitine (63%) and not for the combination therapy (37%). For the secondary outcomes of fatigue, propionyl-L-carnitine and the combination therapy showed a reduction on the 20-point general fatigue axis (from 18.4 SD 1.8 to 16.5 SD 3.1, and from 19.1 SD 1.4 to 17.3 SD 3.3 respectively), whereas acetyl-L-carnitine showed a reduction on the 20-point mental fatigue axis (from 16.3 SD 2.5 to 13.9 SD 3.5). No differences were found on the physical fatigue axis. Patients reported sleeplessness and feeling overstimulated although withdrawal due to harms were similar between groups.

A poor-quality crossover trial randomized patients to an extract of pollen (antioxidant) or placebo for 3 months followed by a 2-week washout and then to the pollen extract or placebo for an additional 3 months resulting in five people in placebo/pollen extract arm, five people in pollen extract/placebo arm, six people in placebo/placebo arm, and six people in pollen

extract/pollen extract arm. They measured total well-being, fatigue, and fatigability on a 10-point Likert scale (0 =no problem and 10=serious symptoms).[100] They found no difference in any of their measures but did note that in the pollen extract group, 62 percent (13/21) reported improvement compared with 23 percent (5/22) in the placebo group. Adherence was not reported and no serious harms were noted.

A small fair-quality trial (n=86), randomized patients diagnosed with CFS based on the Oxford (Sharpe, 1991) criteria to homeopathy or placebo.[104] Homeopathic prescriptions included different single or multiple remedies prescribed at each consultation over a 6-month period and found improvement on the general fatigue subscale of the MFI-20 (mean change 2.70, SD 3.93 vs. 1.35, SD 2.66, p=0.04) but no difference on other dimensions or on the Fatigue Impact Scale and in the proportion of patients clinically improving (a change from baseline of 15% in MFI subscales). They did report improvement on the physical dimension subscale of the Functional Limitations Profile (FLP) with a mean change of 5.11 (SD 8.82) compared with 2.72 (SD 8.40) in the placebo arm. Attrition was similar between groups (overall 11/103, 11%) and neither adherence nor harms were reported.

One small (n=39) fair-quality trial compared a low sugar/low yeast diet with healthy eating in a group of primarily female patients (88%) diagnosed with ME/CFS using the CDC (Fukuda, 1994) criteria.[99] The low sugar/low yeast diet involved omission of all sugar containing foods, refined carbohydrates, yeast containing foods, alcohol, and caffeine with a limited consumption of fruit and milk except a daily yogurt. Those randomized to the healthy eating approach were advised to consume a high fiber diet with five servings of fruit and vegetables per day, two servings of fish per week, and reduced fat and refined carbohydrate. Patients were followed for 24 weeks for outcomes of fatigue and quality of life. They found no difference in either outcome based on the Chalder Fatigue Scale and the SF-36 but did note high loss to followup (25% and these were not included in analysis) and low adherence (24% in the low sugar/low yeast group vs. 67% in the healthy eating group).

A large good-quality trial (n=409) randomized patients to distant healing versus usual care (waiting) and used the CDC (Fukuda, 1994) or Oxford (Sharpe, 1991) definitions for inclusion.[103] The median duration of illness ranged from 9.6 to 11.9 years. They measured adherence of the healers who were from 21 European countries with a mean healing experience of 9.7 years (SD 7.9 years), and replaced those not complying with the study design (7%, 34/462). In addition to other outcomes not included in this report, they considered the Physical Health Component Summary score of the SF-36 and found no differences (1.11; 95% CI, -0.255 to 2.473). Although there was no interaction effect for treatment and blinding (p=0.32), patients who knew they were not being treated had much lower scores at followup (mean difference between groups: -1.544; 95% CI, -2.913 to -0.176).[103]

A final fair-quality crossover study compared melatonin with phototherapy. Thirty patients identified using the Oxford criteria were given placebo initially for 12 weeks followed by melatonin (5 mg every evening) or phototherapy (2500 Lux for 1 hour in the morning). This was followed by a 12-week washout (phototherapy group) or placebo (melatonin group) and then a crossover to the reverse schedule.[105] They reported no differences in a 10-point visual analogue scale of fatigue, the Mental Fatigue Inventory, or the SF-36 physical functioning dimension.

In summary, there is insufficient evidence to determine the effectiveness or harms of CAM interventions due to small single studies and methodological limitations.

Table 4. Trials of complementary and alternative medicine therapies for ME/CFS

Author, year Study type N Quality	Case definition	Duration/ followup	Interventions	Overall effect Intervention A vs. intervention B vs. intervention C etc.
Hobday, *et al.,* 2008[99] N=39 Fair	CDC (Fukuda, 1994) criteria	6 months (24 weeks)	A. Low sugar/low yeast B. Healthy eating	**Fatigue outcomes:** Chalder Fatigue Scale scores: NS SF-36 vitality subscale scores: NS **Function outcomes:** SF-36 physical functioning subscale scores: NS
Ockerman, 2000[100] N=22 Poor	CDC (Fukuda, 1994) criteria	3 months	A. Pollen: Antioxidant extract of pollen (Polbax) B. Placebo	**Fatigue outcomes:** Fatigue scores improved in the pollen group on the Likert at 3 months: -0.43 vs. -0.18; p<0.05 **Quality of life outcomes:** Total well-being scores lower in the placebo group at 3 months: 7.14 vs. 6.66; p=NR Change from baseline in the pollen group vs. placebo: -1.66 vs. -0.21; p<0.01 Change in total well-being after treatment: p=NR Worse: 9.5% (2/21) vs. 18% (4/22) No change: 29% (6/21) vs. 59% (13/22) Better: 62% (13/21) vs. 23% (5/22)
The, *et al.,* 2007[101] N=57 Good	CDC (Fukuda, 1994) criteria	3.5 months (14 weeks)	A. Acclydine B. Placebo	**Fatigue outcomes:** CIS fatigue severity scores: NS **Function outcomes:** SIP-8 scores: NS **Other outcomes:** Physical activity level over a 12-day period: NS
Vermeulen and Scholte, 2004[102] N=89 Fair	CDC (Fukuda, 1994) criteria	6 months (24 weeks)	A. Acetyl-L-carnitine (ALC) B. Propionyl-L-carnitine (PLC) C. Combination, Acetyl-L-carnitine 2 g/day + propionyl-L-carnitine 2 g/day (combo)	**Fatigue outcomes:** General fatigue scores better in ALC group based on MFI-20 scores at 24 weeks, mean (SD): 15.9 (4.2) vs. 16.5 (3.1) vs. 17.3 (3.3) However, both other interventions improved more from baseline: p=0.004 for PLC change from baseline; p=0.000 for combo change from baseline Mental fatigue at 24 weeks: 15.1 (3.6) vs. 13.9 (3.5) vs. 14.6 (4.0); p=0.015 for ALC change from baseline **Other outcomes:** % Improved from baseline was higher in the PLC group on CGI at 24 weeks: 59% (17/29) vs. 63% (16/unclear) vs. 37% (11/30)

Author, year Study type N Quality	Case definition	Duration/ followup	Interventions	Overall effect Intervention A vs. intervention B vs. intervention C etc.
Walach, *et al.,* 2008[103] N=409 Good	CDC (Fukuda, 1994) criteria	6 months treatment Followup to 18 months	A. Distant healing B. Usual care	**Function outcomes:** SF-36 physical functioning: NS SF-36 physical functioning subscale: NS Covariance analysis effect for blinded vs. unblinded treatment, 95% CI : -1.54 (SE 0.70) -2.91 to -0.18
Weatherly-Jones, *et al.,* 2004[104] N=86 Fair	Oxford (Sharpe, 1991) criteria	6 months	A. Homeopathy B. Placebo	**Fatigue outcomes:** General fatigue scores better in placebo group on MFI-20 scores at 6 months, mean (SD): 2.70 (3.93) vs. 1.35 (2.66), p=0.04 Physical fatigue: NS Mental fatigue: NS FIS Cognitive dimension: NS FIS Physical dimension: NS FIS Social dimension: NS **Function outcomes:** Functional Limitations Profile scores: NS
Williams, *et al.,* 2002[105]N=30 Fair	Oxford (Sharpe, 1991) criteria	12 months (12 weeks treatment, 12 week washout, then 12 week crossover and 12 week washout)	A. Melatonin B. Phototherapy	**Fatigue outcomes:** (IQR) VAS score: NS Mental Fatigue Inventory scores: NS (IQR) SF-36 vitality subscale scores: NS **Function outcomes:** (IQR) SF-36 physical functioning subscale scores: NS

ACL = Acetyl-L-carnitine; CDC= Centers for Disease Control and Prevention; CGI= Clinical Global Impression change score; CI= confidence interval; CIS= Checklist of Individual Strength; FIS= Fatigue Impact Scale; g= gram; IQR= interquartile range; MFI-20= Multidimensional Fatigue Inventory 20-Item; N= sample size; NR= not reported; NS= not significant; PCL= Propionyl-L-carnitine; SD= standard deviation; SE= standard error; SF-36= 36-item Short Form Survey; SIP-8= Sickness Impact Profile 8-Item; VAS= visual analog scale; vs.= versus.

Exercise Programs

One good-quality[98] and five fair-quality randomized trials compared one form of exercise with another form of exercise, standard medical care, adaptive pacing, or placebo (**Table 5** below; **Appendix G4 and H2**).[75,106-109] Studies were conducted in the United Kingdom,[75,98,106,109] United States,[84] New Zealand,[108] and China[107] and enrolled patients based on the CDC (Fukuda, 1994) criteria[84,107-109] or the Oxford (Sharpe, 1991) criteria.[75,98,106] All were intended to target symptoms of ME/CFS. Major limitations of studies include dissimilarity of groups,[106,108] high loss to followup,[109] lack of intention-to-treat analysis of outcomes,[109] and unclear or inadequate blinding.[75,98,106-108] One trial was funded by a ME/CFS network and all other trials were funded by research agencies or trusts.

There is low strength of evidence that exercise therapy was superior to control groups in measures of fatigue, function, and clinical impression of change. There is insufficient evidence based on one small study to determine the effectiveness of home orthostatic training or harms of exercise therapies.

The effectiveness of GET compared with control groups (usual care, placebo, placebo exercise, adaptive pacing) was studied in four trials.[75,98,106,108] The largest of these was the PACE trial, a 12 month good-quality 4-arm trial that included a comparison of GET (n=159) with adaptive pacing (n=159) and usual care (n=157), and measured outcomes of fatigue (Chalder Fatigue Scale), function (SF-36 physical function subscale), clinical global impression of change, and work impairment (work and social adjustment scale).[98] The CBT arm is discussed in the Cognitive and Behavioral Therapies and Head-to-Head/Comparison Trials Sections. GET consisted of a maximum of 15 sessions including education and a negotiated exercise plan with incremental activity increases aimed at 30 minutes of light exercise 5 times a week. Adaptive pacing consisted of a maximum of 15 sessions of therapy aimed at achieving optimum adaptation to the illness through activity pacing and advice to avoid activities that demand more than 70 percent of participant's perceived energy. Compared with the usual care group and the adaptive pacing therapy group, at 1 year the GET groups reported statistically significantly better fatigue scores (mean difference GET vs. usual care: -3.2; 95% CI, -4.8 to -1.7; p=0.0003; GET vs. adaptive pacing therapy: -2.5; 95% CI, -4.2 to -0.9; p=0.0059), functioning scores (mean difference GET vs. usual care: 9.4; 95% CI, 4.4 to 14.4; p=0.0005; GET vs. adaptive 12.8; 95% CI, 7.7 to 17.9; p<0.0001), and work impairment scores (mean score: 20.5 GET vs. 23.9 usual care, p<0.001; GET vs. 24.5 adaptive pacing therapy, p<0.001). GET groups reported greater improvement on the self-rated CGI at 1 year compared with both the usual care and adaptive pacing therapy groups (OR of positive change vs. negative change for GET vs. usual care: 2.0; 95% CI, 1.2 to 3.5; p=0.013; GET vs. adaptive pacing therapy: 1.5; 95% CI, 1.0 to 2.3; p=0.028). They also found a significant improvement in PEM in the exercise group compared with both usual care (OR 0.5, p=0.003) and adaptive pacing (OR 0.5, p=0.004). Non-serious harms were reported often and were similar between groups (usual care: 977; adaptive pacing therapy: 949; GET: 992). Although serious hars and serious clinical deterioration were uncommon, there were more reported in the exercise group (17) than the usual care group (7), p=0.04. They also considered serious adverse reactions to the treatment intervention and found no differences between groups.

A smaller fair-quality trial of shorter duration found similar results on outcomes of fatigue and CGI, but did not find improvement in the physical function subscale of the SF-36. This was a 12-week trial with 6-month followup comparing GET with standard medical care (n=49).[108] Exercise consisted of treadmill walking starting at 10 to 15 minutes, 4 to 5 times per week at a

heart rate of 40 percent of VO_2 max (50% maximal heart rate) and increased by 3 to 5 minutes per week for the first 6 weeks and then by an increase in heart rate by 5 beats per minute per week with a goal of achieving 30 minutes of exercise at 70 percent VO_2 maximum (80% maximal heart rate) at 12 weeks.[108] The primary outcome was the CGI, which was significantly improved in the GET group compared with the standard medical care group at 12 weeks (55% rated as much or very much better compared with 24% of the control, p=0.04). Sixty-eight percent of patients rated their exercise therapy as 'effective' or 'highly effective'. Compared with standard therapy, GET showed improvement in all of the secondary fatigue outcomes at 12 weeks (Chalder Fatigue Scale scores; total score: -10.54 vs. -0.94, p=0.02; physical fatigue subscale: -6.64 vs. -0.34, p=0.02; mental fatigue subscale: -3.90 vs. 0.60, p=0.03), but showed no difference in the SF-36 physical functioning subscale (15.95 vs. 9.35, p=0.49). In their intention-to-treat analysis with a 12 percent dropout rate (3 per group), neither the Chalder Fatigue Scale mental fatigue subscale or SF-36 physical functioning subscale were significant.[108] Notably, the GET group was younger (mean age 37 vs. 45 years) and had a shorter duration of illness (2.7 vs. 5.0 years). They received 77 percent of the questionnaires at the 6 month followup and found sustained improvement on the CGI and the Chalder Fatigue Scale physical fatigue whereas the Chalder Fatigue Scale mental fatigue subscale scores showed no difference between groups at 6 months.[108] They also considered physiological assessment of fitness with incremental testing on a treadmill to determine maximum aerobic capacity (VO_2 peak) and found no difference between groups, however, complete data was only available for just over half the sample as 10 patients refused to have a second test due to perceived harm from the initial testing, five stopped prior to maximal effort, and the equipment failed on another two patients.

An earlier fair-quality 12-week trial conducted in the United Kingdom found similar results for CGI, measures of fatigue, and function measured by the SF-36 when comparing GET with flexibility exercises.[106] Patients attended weekly sessions in which they were prescribed a home program that consisted of either exercise (primarily walking) or stretching and relaxation to be performed 5 days per week. The initial exercise prescription consisted of 5 to 15 minutes of aerobic exercise at an intensity of 40 percent of peak oxygen consumption (50% maximum heart rate) to be increased by 1 or 2 minutes to a maximum of 30 minutes. Once achieved, the intensity was then increased to a maximum of 60 percent of peak oxygen consumption, as monitored by heart rate. If fatigue increased, patients were advised to maintain the same level of exercise until fatigue lessened. The flexibility and relaxation group started at 10 minutes of stretching/relaxation and were advised to increase to a maximum of 30 minutes while avoiding any extra physical activities. A greater number of patients in the GET group reported "much" or "very much" improvement on the CGI (16/29, 55% vs. 8/30, 27%; p=0.05).[106] Intention-to-treat analysis including seven patients who dropped out (4 exercise, 3 flexibility/relaxation) found similar results (17/33 vs. 9/33, p=0.04) They also evaluated changes in fatigue using various measures and found significant improvement on all measures with the exception of the mental fatigue subscale of the Chalder Fatigue Scale based on differences in means (Chalder Fatigue Scale total score: -8.40 vs. -3,10, p<0.01; VAS total fatigue score [normal=200]: -59 vs. -39, p=0.04; VAS physical fatigue score [normal=100]: -31 vs. -23, p<0.01; VAS mental fatigue score [normal=100]: not significant). Improvement was also noted in function based on SF-36 total scores (137 vs. 84, p=0.05) and physical function score (47.5 vs. 8.0, p=0.01). Although they reported a significant difference in the SF-36 general health score between groups at 12 weeks, there was a difference between the groups at baseline with the change being similar (4.0 vs. 4.0). Differences were also noted in peak oxygen consumption, mean heart rate during

submaximal treadmill testing, and mean submaximal perceived exertion score favoring the exercise group (13% vs. 6%; 143 beats per minute, SD 13 vs. 150 beats per minute, SD 13; 14.5, SD 3.4 vs. 16.2, SD 2.8, respectively). Twenty-three of thirty patients in the flexibility/relaxation group were allowed to crossover to the exercise intervention at 12 weeks. One dropped out due to an unrelated condition and of the 22 who completed the program, 54 percent (12/22) rated themselves as better. At 1 year they found persistent improvement in measures of function with 66 percent (31/47) of those working or studying at least part time compared with only 39 percent (26/66) at baseline (95% CI, 9% to 44%).

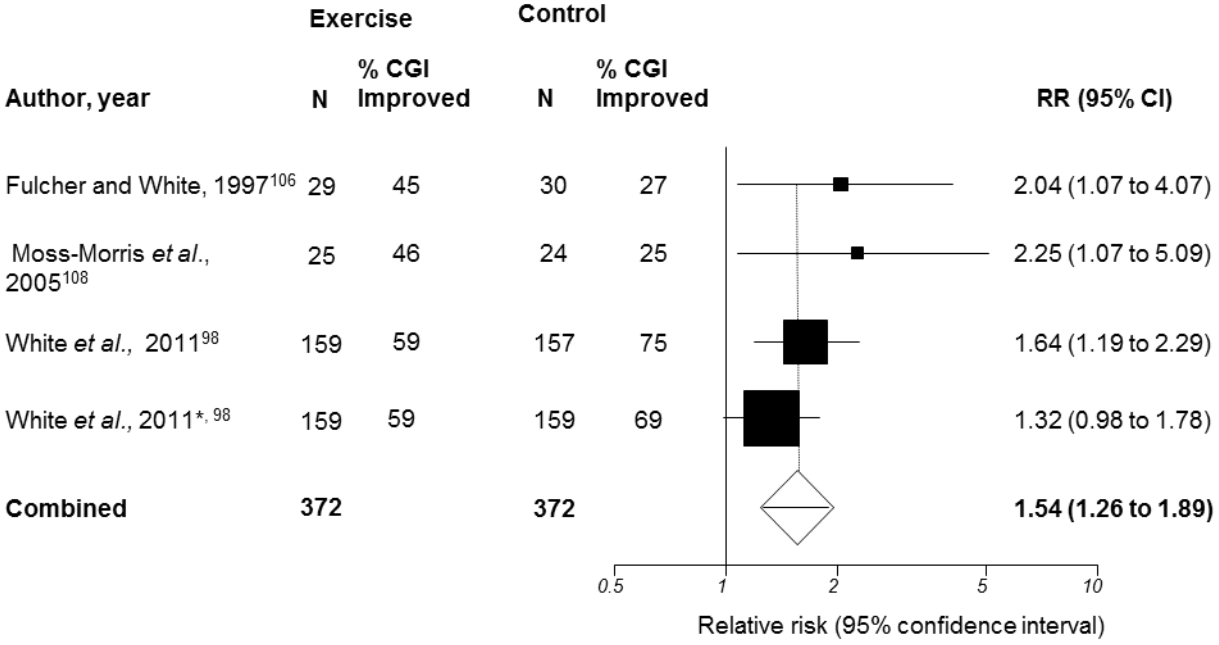
A final fair-quality 4-arm study (n=136) compared fluoxetine with GET, placebo, or a combination of GET and fluoxetine, followed them for 6 months, and measured fatigue using the Chalder Fatigue Scale and functional capacity measuring the amount of oxygen consumed in the final minute of exercise per kg of body weight.[75] Attrition was highest in the combination group (42%), but was also high in the individual intervention groups (32% in fluoxetine group and 29% in GET group), while the control group had lower attrition (17%). Adherence was not reported. Results of the GET versus placebo arms will be reported in this section. After 6 months of treatment there was greater improvement in fatigue and more non-cases of fatigue (Chalder score <4) in the exercise interventions (18% vs. 6%). The exercise interventions showed an improvement in functional capacity, with a mean change from baseline at 6 months of 2.8 ml O_2/kg per minute (95% CI, 0.8 to 4.8). Total withdrawal was greatest in the exercise groups (25/68, 37%) compared with the non-exercise groups (15/69, 22%).

Pooling of three of these studies found a significant improvement in CGI (RR 1.58; 95% CI, 1.25 to 1.98, **Figure 4**), and SF-36 physical function subscale (weighted mean difference 10.29; 95% CI, 6.71 to 13.88, **Figure 5**).[98,106,108] Due to the variability in how the Chalder fatigue scale was used across studies, these results could not be pooled.

Two other forms of exercise were studied and found mixed results. Ho and colleagues performed a 4-month randomized trial of qigong exercise compared with no qigong exercise in 64 patients ages 18 to 55 years recruited through online or newspaper advertising in Hong Kong.[107] Patients randomized to the exercise group received group qigong twice weekly for 5 weeks (2 hours of education, relaxation, stretching, and 1 hour of qigong training per session) followed by 12 weeks of home based qigong (30 minutes per day). The control group was asked to refrain from qigong exercise. Attrition was 19 percent and similar between groups. Improvement was reported in the Chalder Fatigue Scale total score (mean change: -18.3 vs. -7.6, $p<0.001$), physical fatigue subscale score (mean change: -12.1 vs. -4.4, $p<0.001$), and mental fatigue subscale score (mean change: -5.1 vs. -3.1, $p=0.01$) in both groups with a significantly greater improvement in the qigong exercise group. No change was noted in the SF-36 physical function subscale score (3.2 vs.2.1, $p=0.48$). Home orthostatic training (40 minutes of standing against a wall with their heels 15 cm from the wall) was compared with a sham home exercise program (10 minutes of wall standing while performing intermittent calf contractions) in a fair-quality 6-month trial of 38 patients with CFS based on the CDC (Fukuda, 1994) criteria.[109] No differences in fatigue as measured by the fatigue impact score in those who completed the trial and submitted a final questionnaire (n=25). At 6 months, the sham group had a significantly greater drop in blood pressure when standing compared with the intervention group (-6 mmHg, 95% CI, 0.0 to 12.6, $p=0.05$), but the clinical significance of this was not reported. Of note, they did not perform any subgroup analysis to determine if differences existed in those with subjective autonomic symptoms at baseline.

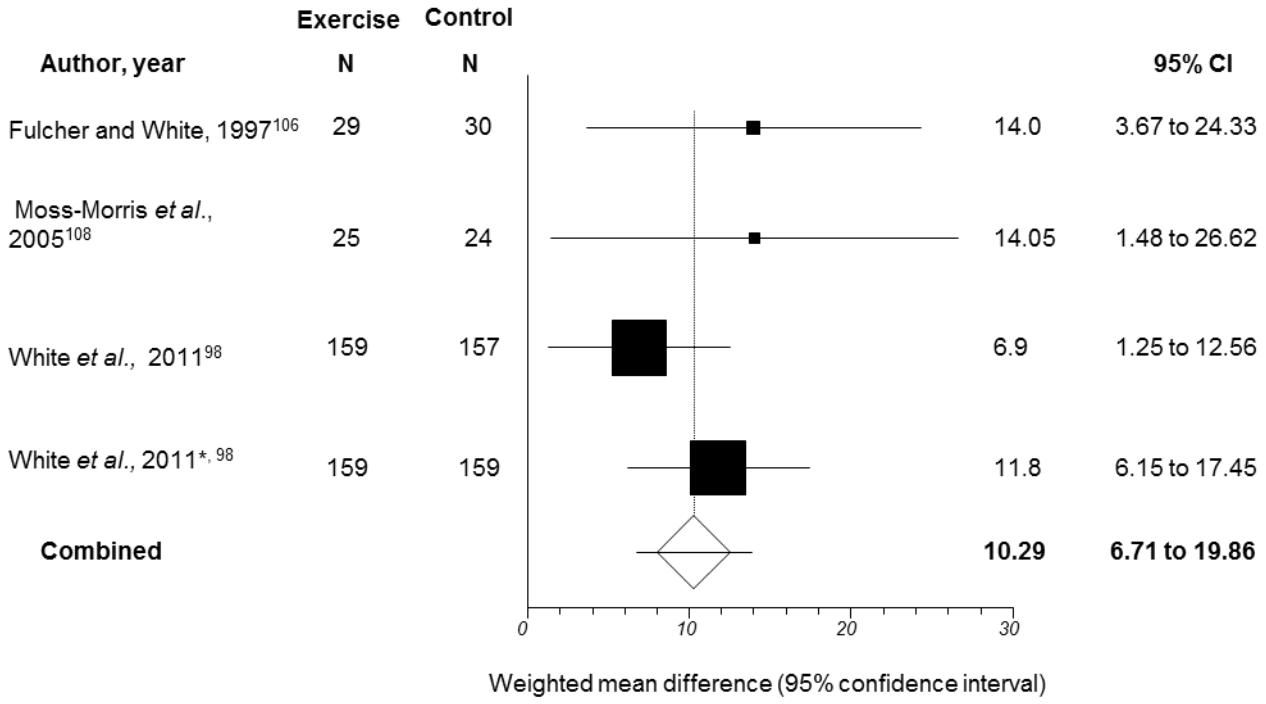
In summary, GET improves function (moderate strength), and global improvement (moderate strength), and fatigue (low strength) in ME/CFS patients compared with control groups. Although qigong exercise found improvement in measures of fatigue and orthostatic training found no different in measures of fatigue but did improve physiological measures, this represents an insufficient evidence based on single small studies. Harms were not well reported but the high rate of patients refusing repeat exercise testing in one study due to concern of harm suggests that this outcome has not been adequately studied.

Figure 4. Meta-analysis of improvement on CGI scale for exercise compared with controls



*Using adaptive pacing as the comparison
Abbreviations: CGI = Clinical Global Impression of Change Score; CI= confidence interval; N= sample size; RR= relative risk.

Figure 5. Meta-analysis of mean changes in SF-36 physical function subscale scores for exercise compared with controls



*Using adaptive pacing as the comparison
Abbreviations: CI= confidence interval; N= sample size; RR= relative risk; SF-36= Short Form 36-item survey.

Table 5. Trials of exercise programs for ME/CFS

Author, year Study type N Quality	Case definition	Duration/ followup	Interventions	Overall effect Intervention A vs. intervention B vs. intervention C etc.
Fulcher and White, 1997[106] RCT N=59 Fair	Oxford (Sharpe, 1991) criteria	3 months (12 weeks), 1 year followup	A: Exercise group B: Control group	**Fatigue outcomes:** Chalder Fatigue Scale mean (SD): 20.5 (8.9) vs. 27.4 (7.4); p=0.004 Physical fatigue scores better mean (SD): 130 (28) vs. 154 (34); p=0.006 Other VAS fatigue scores: NS **Function outcomes:** SF-36 physical functioning subscale mean (SD): 69 (18.5) vs. 55 (21.8); p=0.01 **Employment outcomes:** working full- or part-time at 1 year followup: 66% (31/47) vs. 39% (26/66); 95% CI, 9% to 44% **Other outcomes:** Self-rated CGI scores of "very much better" :: 31% (9/29) vs.7% (2/30), p=0.05 Median (IQR) peak O_2 consumption: NS Median increase in peak O_2 consumption: NS Median increase in isometric strength: NS
Ho, *et al.,* 2012[107] RCT N= 52 Fair	CDC (Fukuda, 1994) criteria	4 months	A: Qigong exercise B: Control group	**Fatigue outcomes:** Chalder Fatigue Scale mean (SD): 21.6 (10.4) vs. 32.1 (8.8); p=0.000 Physical fatigue scores, Chalder Fatigue Scale mean (SD): 12.9 (6.1) vs. 20.3 (5.7); p=0.000 Mental fatigue scores on Chalder Fatigue Scale mean (SD): 8.8 (4.6) vs. 11.9 (3.8); p=0.012 **Function outcomes:** SF-12 mental functioning subscale mean (SD): 42.7 (7.2) vs. 35.7 (9.5); p=0.001 SF-12 physical functioning subscale: NS **Other outcomes:** Telomerase activity at 4 months (arbitrary unit): 0.178 (0.201) vs. 0.104 (0.059); p=0.029

Author, year Study type N Quality	Case definition	Duration/ followup	Interventions	Overall effect Intervention A vs. intervention B vs. intervention C etc.
Moss-Morris, *et al.,* 2005[108] RCT N= 49 Fair	CDC (Fukuda, 1994) criteria	3 months (12 weeks), 6 month followup	A: Exercise group B: Control group	**Fatigue outcomes:** Chalder Fatigue Scale mean (SD): 13.91 (10.88) vs. 24.41 (9.69); p=0.02 Physical fatigue scores on Chalder Fatigue Scale mean (SD): 7.91 (7.06) vs. 14.27 (5.75); p=0.02 Mental fatigue scores on Chalder Fatigue Scale mean (SD): 6.00 (4.06) vs. 10.14 (4.27); p=0.03 **Function outcomes:** SF-36 physical functioning: NS **Other outcomes:** Self-rated CGI scores of % “much or very much improved" at 6 months: 54 (12/22) vs. 23.8 (5/21); p=0.04
Sutcliffe, *et al.,* 2010[109] RCT N= 36 Fair	CDC (Fukuda, 1994) criteria	6 months	A. Orthostatic training B. Control group	**Fatigue outcomes:** Mean FIS scores: NS **Function outcomes:** Mean systolic blood pressure: NS Mean heart rate (beats per minute): NS
Wearden, *et al.,* 1998[75] N=68 Fair *Same study as on Table 3 and 6*	Oxford (Sharpe, 1991) criteria	6.5 months (26 weeks)	A. GET B. Placebo control*	**Fatigue outcomes:** Fatigue scores significantly improved in GET group on Chalder Fatigue Scale at 26 weeks, mean change from baseline (95% CI): -5.7 (-9.5 to -1.9) vs. -2.7 (-5.4 to 0.01) Non-cases of fatigue on Chalder Fatigue Scale (score <4) at 26 weeks: 18% (6/33) vs. 6% (2/34), p=0.025 for exercise interventions combined vs. control **Function outcomes:** Functional work capacity based on amount of O_2 consumed in the final minute of exercise per kg of body weight improved in GET group at 26 weeks, mean change (95% CI) : 2.8 (0.8 to 4.8) vs. -0.1 (-1.7 to 1.6) Effect of exercise on functional work capacity, mean change 0-26 weeks: 1.9 (95% CI 0.15 to 3.69), p=0.03

Author, year Study type N Quality	Case definition	Duration/ followup	Interventions	Overall effect Intervention A vs. intervention B vs. intervention C etc.
White, *et al.,* 2011[98] ***PACE Trial*** N=479 Good *Same as on Table 3 and 6*	Oxford (Sharpe, 1991) criteria	13 months (52 weeks)	A. APT B. GET C. Usual care[†]	**Fatigue outcomes:** Fatigue scores better in GET group on Chalder Fatigue Scale at 52 weeks, mean (SD): 20.6 (7.5) vs. 23.1 (7.3) vs. 23.8 (6.6) Mean difference GET vs. APT at 52 weeks (95% CI): -2.5 (-4.2 to -0.9) p=0.0059 Mean difference GET vs. usual care at 52 weeks (95% CI):. -3.2 (-4.8 to -1.7) p=0.0003 % Improved from baseline (by ≥2 points) higher in GET group: 80% (123/154) vs. 65% (99/153) vs. 65% (98/152) **Function outcomes:** Functional scores better in GET group on SF-36 physical functioning subscale at 52 weeks, mean (SD): 57.7 (26.5) vs. 45.9 (24.9) vs 50.8 (24.7) Mean difference GET vs. APT at 52 weeks (95% CI): 12.8 (7.7 to 17.9) p<0.0001 Mean difference GET vs. usual care at 52 weeks (95% CI) : 9.4 (4.4 to 14.4) p=0.0005 % Improved from baseline (by ≥8 points) higher in GET group: 70% (108/154) vs. 49% (75/153) vs. 58% (88/152) **Employment outcomes:** Work and social adjustment scale scores better in GET group at 52 weeks, mean (SD): 20.5 (9.4) vs. 24.5 (8.8) vs. 23.9 (9.2); p=0.0006 for GET vs. control and p=0.0004 for GET vs. APT **Other outcomes:** More with positive change in GET group on self-rated CGI at 52 weeks: 41% (62/152) vs. 31% (47/153) vs. 25% (38/152) Minimum change: 53% (80/152) vs. 63% (96/153) vs. 66% (100/152) Negative change: 7% (10/152) vs. 7% (10/153) vs. 9% (14/152) *The positive change vs. negative change, OR (95% CI)* GET vs. APT: 1.5 (1.0 to 2.3) p=0.028 GET vs. usual care: 2.0 (1.2 to 3.5) p=0.013

*Comparisons between Fluoxetine and placebo are presented in Table 2 and comparisons with GET + fluoxetine are presented in Table 6.

[†]Comparisons for CBT with APT and usual care are presented in Table 3 and comparisons for CBT with GET are presented in Table 6.

APT= adaptive pacing therapy; CDC= Centers for Disease control and Prevention; CGI= Clinical Global Impression Change Score; CI= confidence interval; FIS= Fatigue Impact Scale; IQR= interquartile range; N= sample size; NR= not reported; NS= not significant; O_2= oxygen; RCT= randomized controlled trial; SD= standard deviation; SF-36= 36-item Short Form Survey; SF-12= Short Form 12-item Survey; VAS= visual analog scale; vs.= versus.

Head-to-Head Comparisons and Combination Therapies

Five trials (7 publications) were included that compared either head-to-head interventions or combinations of two interventions (**Table 6** below; **Appendix G4**).[75,77,81,82,84,87,98] Four trials (in 6 publications) compared head-to-head interventions, one of face-to-face CBT compared with telephone CBT,[77] one comparing CBT with GET and adaptive pacing and a usual care group,[98] one comparing CBT with cognitive therapy and anaerobic activity therapy and relaxation,[81,82,84] and one of fluoxetine compared with GET,[75] which also had a group of the combination of fluoxetine plus GET for comparison. The fifth trial compared a combination treatment of CBT plus GET with usual care.[75,87] The results for the head-to-head- trials that pertain to the comparison of one intervention to another, not a control group, on outcomes will be discussed in this section; results based on comparisons with control groups, such as GET or CBT were discussed in those previous sections. All trials were designed to target symptoms of ME/CFS, not treat the underlying cause. One trial was good-quality[98], while the other four were rated fair-quality[75,77,84,87] (**Appendix H2**). Three of the trials were of large size (n>100) and illness duration, when reported, was at least greater than 6 months. Two trials used the Oxford (Sharpe, 1991) criteria to identify patients with CFS,[75,98] one used the CDC (Fukuda, 1994) criteria,[87] while another used a combination of both the Oxford (Sharpe, 1991) criteria and CDC (Fukuda, 1994) criteria,[77] and the other used a combination of a CFS symptom questionnaire, psychiatric assessment, and medical assessment.[84]

Adherence was only reported in two trials, with one reporting participants completed an average of 10 out of 13 sessions[84] and another reporting an average of 11.3 out of 15 sessions.[77] This trial also reported that 20 percent of the face-to-face CBT group and 30 percent of the telephone CBT group did not receive the intervention. Attrition was relatively high (<20%) in three trials, with the telephone CBT group reporting the highest attrition of all (56%), even though attrition in the face-to-face CBT group was also high (34%).[77] The trial of fluoxetine and GET reported the second highest attrition of these trials in the combination group (42%) with both the fluoxetine only group and the GET only groups reporting high attritions (32% vs. 29%), while the no treatment group did not (17%).[75] The other two trials reported low attrition rates (0.6-5%).[87,98]

There is low strength of evidence that GET and CBT or cognitive therapy had similar results on measures of fatigue and function in one good-quality and two fair-quality head-to-head trials. GET was superior to fluoxetine on measures of fatigue and function in one fair-quality trial but represents an insufficient strength of evidence given that it was a single study of small sample size. CBT delivered either face-to-face or over the telephone showed improvement in overall functioning and work impairment in one fair-quality trial, but there were no differences between groups and strength of evidence is insufficient given a single study of small sample size. harms were not well reported leaving insufficient evidence on the harms of GET although patients receiving GET reported more serious harms compared with CBT, adaptive pacing, or usual care in one good-quality trial.

The PACE trial described previously was a large 12-month good-quality trial (n=641) comparing four interventions: CBT; GET; an adaptive pacing therapy; and a usual care control group.[98] Attrition was low with only 1.7 percent withdrawing overall and adherence was not reported. Compared with the control and adaptive pacing groups, at 1 year the CBT and GET groups reported similar improvement in fatigue scores, functioning scores, and work impairment scores. Both CBT and GET groups reported great improvement on the self-rated CGI at 1 year

compared with both the control and adaptive pacing therapy groups (OR of positive change vs. negative change for CBT vs. control: 2.2; 95% CI 1.2 to 3.9; p=0.011; CBT vs. adaptive pacing therapy: 1.7; 95% CI, 1.0 to 2.7; p=0.034; GET vs. control: 2.0; 95% CI, 1.2 to 3.5; p-0.013; GET vs. adaptive pacing therapy: 1.5; 95% CI, 1.0 to 2.3; p-0.028). Harms were reported often, however the CBT group reported fewer events (848) compared with all groups (control: 977; adaptive pacing therapy: 949; GET: 992). Serious harms were rare, but more were reported by the GET group (17) compared with all groups (control: 7; adaptive pacing therapy: 16; CBT: 8).

One fair-quality trial (n=114) compared CBT with cognitive therapy, with an anaerobic activity therapy, and with a relaxation techniques control group, followed them for 12 months and measured fatigue using the FSS, quality of life using the QLS, functioning using the SF-36 physical functioning subscale, and employment status.[84] Overall 25 percent of individuals dropped out, but this was not reported per group, and individuals attended an average of 10 out of 13 sessions. At 12 months the CBT group and the cognitive therapy group had statistically significantly better functioning scores compared with the anaerobic activity therapy group (mean scores: 58.64 vs. 61.09 vs. 39.72; p<0.01). There were no differences in employment status, fatigue scores, or quality of life scores, and harms were not reported. Two additional publications performed subgroup analyses on this trial population. One article analyzed fatigue and functioning outcomes based on whether or not individuals stayed within their energy envelope, meaning they avoided overexertion and under exertion by exerting a comfortable range of energy, or strayed outside their energy envelope.[82] Individuals rated their perceived energy and expended energy and this was used to determine which individuals stayed within their energy envelope (n=49) and which were outside their energy envelope (n=32). At 12 months there was a statistically significant improvement in mean fatigue and functioning scores from baseline for those who stayed within their energy envelope compared with those who were outside their energy envelope (fatigue scores: -0.9 vs. 0.1; p<0.01 and functioning scores: 17 vs. 0; p=0.03). The second additional analysis article compared fatigue and functioning outcomes based on homework compliance.[81] They identified three groups based on the amount of homework completed; minimum compliance completed 0 to 25 percent, moderate compliance completed 25.1 to 75 percent, and maximum compliance completed 75.1 to 100 percent of their assigned homework. When they assigned individuals to groups they noted that the highest percentage in the maximum group (56%) were in the cognitive therapy group, the highest percentage in the moderate group (34%) were in the CBT group, and the highest percentage in the minimum group (38%) were in the anaerobic and relaxation groups. At 12 months, though there was a trend toward better improvement in fatigue and functioning scores for the maximum compliance group compared with the other groups this did not reach significance.

One fair-quality trial (n=80) compared face-to-face CBT with telephone CBT,[77] followed them for 12 months, and measured fatigue using the Chalder Fatigue Scale, functioning using the SF-36 physical functioning subscale, work impairment using the Work and Social Adjustment scale, and overall improvement using a self-rated global improvement Likert style scale, similar to the CGI scale (ranging from very much worse to very much better). There were no significant differences between groups on any of the outcomes at any time point. Both groups showed significant improvement at 12 months after the end of treatment from baseline on the SF-36 physical functioning subscale and the Work and Social Adjustment scale, and the majority of participants rated themselves as much better or very much better on the self-rated global improvement scale (56%).

Fluoxetine and/or GET were compared in a previously described trial with 6-month measurements of fatigue (Chalder Fatigue Scale) and functional capacity measuring the amount of oxygen consumed in the final minute of exercise per kg of body weight.[75] Exercise was superior to fluoxetine and/or placebo on all measures. Total withdrawal was greatest in the exercise group compared with the non-exercise group (37% vs. 22%). The other combination trial was a fair-quality trial (n=120) that compared group CBT and group GET with a usual care control group, followed them for 12 months, and measured fatigue using the FIS and function using the SF-36 physical functioning subscale.[87] Adherence was not reported, but attrition was low overall (4.2%). Neither fatigue nor functioning scores were significantly different at 12 months followup and harms were not reported.

In summary, head-to-head trials had mixed results with two trials finding improvement with GET, two trials finding improvement with CBT, and one trial finding no differences between CBT, GET, and usual care. In considering non-head-to-head trial data, there is low strength evidence that CBT and GET provide similar improvement in measures of fatigue and/or functioning. It appears that GET is superior to fluoxetine but the strength of evidence is insufficient given that this comparison was only studied in one small fair-quality trial. Although specific side effects were not reported, number of events were greater in the GET groups in two trials.

Table 6. Trials of combination therapies for ME/CFS

Author, year Study type N Quality	Case definition	Duration/ followup	Interventions	Overall effect Intervention A vs. intervention B vs. intervention C etc.
Burgess, *et al.,* 2012[77] N=43 Fair	CDC (Fukuda, 1994) and Oxford (Sharpe, 1991) criteria	12 months	A. Face-to-face B. Telephone	**Fatigue outcomes:** Chalder Fatigue Scale scores: NS **Function outcomes:** SF-36 physical functioning subscale: NS for between groups; p=0.043 for change from baseline for both groups **Employment outcomes:** Work and social adjustment scale scores: NS for between groups; p=0.013 for change from baseline for both groups **Other outcomes:** Global improvement score of "much better or very much better" at 6 and 12 months were higher in the face-to-face group: 60% (15/25) vs. 40% (8/20), 12 months: 57% (13/23) vs. 55% (11/20)' p=NR
Jason, *et al.,* 2007[84] Jason, *et al.,* 2009[82] Hlavaty, *et al.,* 2011[81] N=114 Fair *Same study as on Table 3*	CFS Questionnaire, psychiatric assessment for DSM-IV diagnosis, and medical assessment	12 months	A: CBT B: COG C: ACT D. Relaxation	**Fatigue outcomes:** FSS scores mean (SD): 5.37 (1.19) vs. 5.87 (1.01) vs. 5.77 (1.43) vs. 5.62 (1.06); p=NR Comparison by energy envelope for those who stayed within envelope vs. outside envelope at 12 months was 5.3 vs. 6.3. Change at 12 months from baseline: -0.9 vs. 0.1; p<0.01 The comparison by homework compliance level, change in score at 12 months from baseline: -0.17 (0.73) vs. -0.51 (1.00) vs. -0.54 (1.09); p=NR **Quality of life outcomes:** QLS scores mean (SD): 69.10 (18.99) vs. 72.52 (10.84) vs. 63.00 (13.86) vs. 72.00 (19.70); p=NR **Function outcomes:** SF-36 physical functioning subscale scores mean (SD): 58.64 (30.44) vs. 61.09 (23.74) vs. 39.72 (27.63) vs. 61.20 (27.70) p<0.01, for CBT and COG over time vs. ACT over time Comparison by energy envelope for those who stayed within envelope vs. outside envelope at 12 months: 65 vs. 43, change at 12 months from baseline: 17 vs. 0; p=0.03 Comparison by homework compliance level, change in score at 12 months from baseline: 6.99 (19.30) vs. 7.55 (18.85) vs. 17.50 (18.09); p=NR % Achieving clinically significant improvement: NS **Employment outcomes:** % Employed at 12 month followup: NS

Author, year Study type N Quality	Case definition	Duration/ followup	Interventions	Overall effect Intervention A vs. intervention B vs. intervention C etc.
Núñez, *et al.,* 2011[87] N=115 Fair	CDC (Fukuda, 1994) criteria	2.5-3 months of treatment, 12 months followup after treatment	A. CBT + GET B. Usual care	**Fatigue outcomes:** FIS score for CBT + GET vs. usual care: NS **Function outcomes:** SF-36 physical function subscale for CBT + GET vs. usual care: NS
Wearden, *et al.,* 1998[75] N=136 Fair	Oxford (Sharpe, 1991) criteria	6.5 months (26 weeks)	A. GET + fluoxetine B. GET + drug placebo C. Fluoxetine + exercise placebo D. Placebo control	**Fatigue outcomes:** Chalder Fatigue Scale a, mean change from baseline (95% CI): -6.0 (-9.7 to -2.3) vs. -5.7 (-9.5 to -1.9) vs. -3 (-5.9 to -0.2) vs. -2.7 (-5.4 to 0.01) Non-cases of fatigue on Chalder Fatigue Scale (score <4) at 26 weeks: 18% (6/33) vs. 18% (6/34) vs. 6% (2/ 35) vs. 6% (2/34), p=0.025 for exercise interventions combined vs. others Exercise improved fatigue scale scores: NS **Function outcomes:** Functional work capacity based on amount of O_2 consumed in the final minute of exercise per kg of body weight higher in GET group at 26 weeks, mean change (95% CI) : 2.0 (0.4 to 3.5) vs. 2.8 (0.8 to 4.8) vs. 1.0 (-0.9 to 3.0) vs. -0.1 (-1.7 to 1.6) Effect of exercise on functional work capacity, mean change 0-26 weeks: 1.9 (95% CI 0.15 to 3.69), p=0.03
White, *et al.,* 2011[98] ***PACE Trial*** N=630 Good *Same study on Table 3 and 5*	Oxford (Sharpe, 1991) criteria	13 months (52 weeks)	A. APT B. CBT C. GET D. Usual care	**Fatigue outcomes:** Fatigue scores: p=NS for CBT vs. GET % Improved from baseline (by ≥2 points) was similar in CBT and GET groups: 65% (99/153) vs. 76% (113/148) vs. 80% (123/154) vs. 65% (98/152) **Function outcomes:** Functional scores: p=NS for CBT vs. GET % Improved from baseline (by ≥8 points) was similar in the CBT and GET groups: 49% (75/153) vs. 71% (105/148) vs. 70% (108/154) vs. 58% (88/152) **Employment outcomes:** Work and social adjustment scale scores: p=NS for CBT vs. GET **Other outcomes:** More with positive change in CBT and GET groups on self-rated CGI at 52 weeks: 31% (47/153) vs. 41% (61/147) vs. 41% (62/152) vs. 25% (38/152)

ACT=anaerobic activity therapy; APT= adaptive pacing therapy; CBT= cognitive behavioral therapy; CDC= Centers for Disease Control and Prevention; CFS= chronic fatigue syndrome; CGI= Clinical Global Impression of Change; CI= confidence interval; COG= cognitive therapy DSM-IV= Diagnostic and Statistical Manual, fourth edition; FIS= Fatigue Impact Scale; FSS= Fatigue Severity Scale; GET= graded exercise therapy; N= sample size; NR= not reported; NS= not significant; O_2= oxygen; QLS= Quality of Life Scale; SF-36= 36-item Short Form Survey; SD= standard deviation; vs.=versus.

Key Question 2c

What are the characteristics of responders and non-responders to interventions?

Key Points

- Evidence on patient characteristics associate with response/non-response to treatment was insufficient as it was limited to one small fair-quality trial that found those who had lower functional impairment, less fatigue, and less pain at baseline were more likely to improve after group CBT

Detailed Synthesis

One fair-quality trial described above,[76] compared group CBT with a wait list control and conducted a separate analysis to compare the baseline measures of those who improved with group CBT (n=10) and those who did not improve (n=17) at 6 months (**Appendix G4**). Those who improved were more likely to have less functional impairment (1,330 vs. 1,985; p=0.031), less daily self-rated observed fatigue (7.4 vs. 9.7; p=0.023), and less daily self-rated observed pain (4.5 vs. 7.8; p=0.026) compared with those who did not improve with group CBT. Though it did not reach statistical significance, those who improved were more likely to be working more hours per week compared with those who did not improve (10.9 vs. 2.6; p=0.062). There were no differences between those who improved on group CBT and those who did not on baseline measures of age, education, duration of illness, CIS fatigue score, psychological distress, depression, physical attributes, self-efficacy, avoidance of activity, and focusing on bodily symptoms.

No other studies evaluated characteristics of responders and non-responders to interventions.

Chapter 4. Discussion

Key Findings

Twenty-eight studies contributed to our understanding of diagnostic methods, diagnostic accuracy or concordance, and harms associated with diagnosis of ME/CFS. Multiple case definitions have been used to define ME/CFS and those that require the symptoms of PEM and neurological and autonomic manifestations appear to represent a smaller but more impaired subset of the broader population. One tool, the artificial neural network test, was found to have good sensitivity, specificity, and accuracy for diagnosing ME/CFS (95%, 85% and 90% respectively). Another, the SOFA-CFS, and certain SF-36 subscales or combination of subscales shows moderate ability to discriminate between patients with ME/CFS compared with those without the condition. None however have been adequately tested in a large population to determine validity and generalizability. Other tests including serum parameters and cardiopulmonary function and recovery, have been insufficiently tested in broad populations to determine utility. We did not find evidence on how diagnostic tests for ME/CFS vary by subgroups of the population or studies on which related conditions should be ruled out prior to making a ME/CFS diagnosis. Evidence suggests that carrying an ME/CFS diagnosis is associated with perceived stigma, financial instability, difficulty in social interactions and relationships, and a greater risk of receiving a psychiatric diagnosis.

Thirty-six trials contributed to our understanding of the efficacy of interventions to treat ME/CFS. Although most of the pharmacological trials were targeting an underlying pathophysiological dysfunction, most of the other interventions were targeting associated symptoms of the disease. Rintatolimod provided benefit in measures of function compared with placebo. Immune modulators and antivirals suggested potential improvement in symptoms including improved measures of exercise, fatigue, performance, activities of daily living, and reduced use of other medications for relief of ME/CFS symptoms but further studies are required to determine if this is replicable. CBT and GET was found to be beneficial compared with control groups for outcomes of fatigue, function, and clinical global impression of change. CBT was also beneficial for outcomes of quality of life and employment. No differences were found for all other interventions and outcomes, as outcomes were either not reported, the study quality was poor, and/or the sample size was inadequate to provide a useful estimate. Although harms were not well reported across trials, GET was associated with a higher number of reported harms and withdrawal rates in several trials.

The key findings for this review are summarized in the summary of evidence table (**Table 7**, below) and the factors used to determine the overall strength of evidence grades are summarized in **Appendix K**.

Strength of Evidence

Results for major clinical outcomes are summarized in a strength of evidence table (**Appendix K**). We did not summarize the strength of evidence on diagnostic methods (Key Question 1) because the methods for doing so are not yet sufficiently developed to account for the variety of study designs, the lack of diagnostic/gold standard, and the uncertainty around determination of precision for estimates of test performance, the lack of consensus about the case definition for identifying a consistent study population, and the absence of a reference/gold

standard.[29] For intervention trials, major clinical outcomes are those explicitly stated in Key Question 2; identified as important outcomes by members of the TEP because they are most relevant to patients, clinicians, and policymakers. Outcomes of benefit included in the strength of evidence table are overall function, fatigue, quality of life, days spent at work/school, proportion working full- or part-time, and clinical global impression of change. Outcomes of harm included in the strength of evidence table are withdrawals due to harms, rates of harms, total withdrawals, serious and total harms.

The strength of evidence table includes the four required domains: study limitations, directness, consistency, precision, and reporting bias (terms defined in **Appendix F)**.[23] The table summarizes the strength of evidence. Where possible, a quantitative estimation of the effect size was provided. When a quantitative estimate was not possible due to the heterogeneity in measuring outcomes and the small number of studies per intervention, a symbolic representation of effect was included with the symbol, + representing benefit, <> representing no difference, and – representing a negative effect.

Study Limitations

Study limitations is the degree to which the included studies for a given outcome have a high likelihood of adequate protection against bias (i.e., good internal validity), assessed through both study design and study conduct. In general, we ranked study limitations as medium to high for all outcomes given the small sample sizes and methodological limitations.

Directness

Directness has two meanings: 1) evidence links the interventions directly to health outcomes, and 2) evidence compares two or more interventions in head-to-head trials. All trials included in this review linked the evidence directly to health outcomes. Only a few trials included multiple interventions and compared the effective treatments directly to each other.

Consistency

Consistency refers to the degree of similarity in the effect sizes of different studies within an evidence base. Consistency was not applicable for the majority of outcomes given that most included only single studies. Consistency was noted in the two small rintatolimod trials on certain measures of function, on CBT trials on measures of fatigue, quality of life, global improvement, and some measures of employment, and on GET trials on measures of function, fatigue, and global improvement.

Precision

Precision is the degree of certainty surrounding an estimate of effect for specific outcomes. The methodology for determining precision for strength of evidence tables emphasizes the need to include both clinical and statistical considerations. Inability to preform meta-analysis, heterogeneity in outcomes and measurement tools, small sample size without power calculations, and lack of clear parameters reflecting a clinically significant result, left most outcomes imprecise. Although the galantamine trial provided sample size calculations, the variability in dosing with multiple subgroups, meant imprecise results. Precision was found in the CBT trials on function, fatigue, global improvement, and work impairment, and in the GET trials on some measures of function, fatigue, and global improvement.

Strength of Evidence

We qualitatively rated the overall strength of evidence as high, moderate, low, or insufficient for each outcome based on the required domains and other relevant factors. Strength of evidence is high for outcomes with low study limitations, consistency, and adequate precision. The strength of evidence was downgraded to moderate for outcomes with medium study limitations, imprecise estimates, inconsistency between trials. Strength of evidence was ranked low if multiple deficiencies existed. Strength of evidence was moderate for CBT and for GET compared with usual care, support, relaxation or adaptive pacing for outcomes of fatigue (CBT), function (GET), and global improvement (CBT, GET). Strength of evidence was low for CBT on measures of function, quality of life, and employment, for GET on measures of fatigue and work impairment, and for rintatolimod on measures of function. For all other interventions and outcomes, strength of evidence was insufficient because these outcomes were either not reported, the study quality was poor, and/or the sample size was inadequate to provide a useful estimate.

Table 7. Summary of evidence

Key question Outcome	Study design Number of studies (n)*	Findings and direction of effect	Strength of evidence grade
KQ1. What methods are available to clinicians to diagnose ME/CFS and how do the use of these methods vary by patient subgroups?			
Available methods	11 cross-sectional and case control studies (n=1,470)	1 good-quality study conducted with a spectrum of patients that included conditions similar to ME/CFS found that the artificial neural network test using a combination of 24 symptoms had good sensitivity, specificity, and accuracy. In that study, CFS symptoms with greatest single-item accuracy were acute onset of fatigue and sore throat.	Not applicable
Available methods	11 cross-sectional and case control studies (n=1,470)	The SOFA-CFS and certain subscales of the SF-36 may be promising for identification of certain components of the CFS criteria, but require further testing in broader populations.	Not applicable
KQ1a. What are widely accepted diagnostic methods and what conditions are required to be ruled out or excluded before assigning a diagnosis of ME/CFS ?	No studies	No studies evaluated strategies for working up a patient.	Not applicable
KQ1b. What is the accuracy and concordance of diagnostic methods?	9 observational/ descriptive studies (n=1,178)	6 studies found that the symptoms reported by different case definitions varied. In general, populations defined by ME or ME/CFS criteria had more severe symptoms or more functional impairment than those defined by CFS criteria alone.	Not applicable
KQ1c. What harms are associated with diagnosing ME/CFS?			
Psychological harm, including stigma from label	5 observational studies (n=677)	5 studies found that patients with CFS feel stigmatized by their diagnosis in terms of financial stability (1 study), work opportunities (1 study), perceived judgments on their character (1 study), social isolation (2 studies), or interactions with the health care system (3 studies).	Not applicable
Misdiagnosis	1 observational study (n=68)	1 study identified a substantial burden of misdiagnosis among the CFS population.	Not applicable
Risk from diagnostic test	No studies	No studies identified that reported objective risks directly related to the process of conducting a diagnostic test for CFS.	Not applicable

Key question Outcome	Study design Number of studies (n)*	Findings and direction of effect	Strength of evidence grade
Prejudice and stereotyping	2 observational studies (n=246)	2 studies identified prejudice and stereotypes within the medical profession; medical trainees and mental health practitioners make judgments about a patient's condition based on the name it carries (ME, CFS, or other) and what treatment is being given.	Not applicable
KQ2a. What are the benefits of therapeutic interventions for patients with ME/CFS and how do they vary by patient subgroups?			
Galantamine vs. placebo			
Decreased fatigue and improved quality of life	1 RCT (n=423)	No significant differences between 4 intervention groups and placebo.	Insufficient
Global improvement	1 RCT (n=423)	No significant differences between 4 intervention groups and placebo.	Insufficient
Improved overall function, increased days spent at work/school and proportion working full- or part-time	No studies	No studies.	Insufficient
Hydrocortisone vs. placebo			
Improved overall function, decreased fatigue, and improved quality of life	1 RCT (n=68)	No significant differences between intervention and placebo.	Insufficient
Increased days spent at work/school and proportion working full- or part-time	No studies	No studies.	Insufficient
Hydrocortisone + fludrocortisone vs. placebo			
Improved overall function, decreased fatigue, and improved quality of life	1 RCT (n=80)	No significant differences between intervention and placebo.	Insufficient
Increased days spent at work/school and proportion working full- or part-time	No studies	No studies.	Insufficient
Immunoglobulin G vs. placebo			
Improved overall function	1 RCT (n=28)	Significantly better scores on SF-36 social functioning scale after intervention compared with placebo ($p<0.05$), but no difference on physical functioning scale.	Insufficient

Key question Outcome	Study design Number of studies (n)*	Findings and direction of effect	Strength of evidence grade
Improved fatigue and quality of life, increased days spent at work/school and proportion working full- or part-time	No studies	No studies.	Insufficient
Rintatolimod vs. placebo			
Improved overall function	1 RCT (n=84)	Significant increase in activities of daily living after intervention compared with placebo (23% vs. 14%, p=0.034), but no difference in change in KPS scores from baseline.	Insufficient
Increased exercise work capacity	2 RCT (n=316)	The intervention group compared with placebo had significant increases in exercise duration (10% vs. 2%, p=0.007), exercise work (12% vs. 6%, p=0.011), and cardiopulmonary exercise tolerance (37% vs. 15%, p=0.047).	Low
Improved quality of life, increased days spent at work/school and proportion working full-or part-time	No studies	No studies.	Insufficient
Valganciclovir vs. placebo			
Decreased fatigue	1 RCT (n=30)	Significant decrease in fatigue based on FSS scores decreasing in intervention group compared with placebo (mean change from baseline: -0.06 vs. 0.02, p=0.006).	Insufficient
Improved overall function	1 RCT (n=30)	No significant differences between intervention and placebo.	Insufficient
Improved quality of life, increased days spent at work/school and proportion working full-or part-time	No studies	No studies.	Insufficient
Isoprinosine vs. placebo			
Improved overall function and decreased fatigue	1 RCT (n=15)	No significant differences between intervention and placebo.	Insufficient
Improved quality of life, increased days spent at work/school and proportion working full-or part-time	No studies	No studies.	Insufficient
Fluoxetine vs. placebo			
Improved overall function	1 RCT (n=68)	No significant differences between intervention and placebo.	Insufficient
Decreased fatigue	1 RCT (n=68)	No significant differences between intervention and placebo.	Insufficient

Key question Outcome	Study design Number of studies (n)*	Findings and direction of effect	Strength of evidence grade
CBT/counseling vs. no treatment or support or relaxation or adaptive pacing			
Improved overall function	12 RCT (n=1,637)	Results were mainly positive, but mixed. When 8 trials using the SF-36 physical functioning subscale were pooled there was a significant effect for the intervention group to have better scores vs. control at followup: weighted mean difference of 7.73 (95% CI 3.58 to 11.87). In 5 trials counseling improved overall functioning vs. controls on various measures (49 to 80% improved in counseling groups vs. 17 to 58% in controls), while 2 trials reported mixed results with different measures in the same study, 1 trial reported improvement in the control group compared with counseling, and the other 4 trials reported no differences between groups.	Low
Decreased fatigue	12 RCT (n=1,635)	Results were primarily positive, but mixed. In 9 trials counseling significantly decreased fatigue vs. controls on various measures (63 to 76% improved in counseling groups vs. 15-65% in controls), while the other 3 trials reported no differences between groups.	Moderate
Improved quality of life	5 RCT (n=539)	Results were mixed. In 2 trials counseling showed an improvement in quality of life vs. controls on various measures (mean QOLS at 12 weeks: 2.81 vs. 3.26; p=0.02 and mean change in QLI scores from baseline at 12 months: 2.6 vs. 0.6; p<0.05), 1 trial reported better quality of life vs. support but not no treatment, and the other 2 trials reported no differences between groups.	Low
Increased proportion working full- or part-time	2 RCT (n=145)	No significant differences between intervention and control.	Low
Increased hours worked	3 RCT (n=321)	Significantly more hours worked per week for CBT group vs. control (mean 35.57 vs. 24.00; p<0.04) for 1 trial. The other 2 trials reported no significant differences between intervention and no intervention.	Low
Decreased work impairment	2 RCT (n=531)	Significant improvement reported in both studies for CBT group on work and social adjustment scale compared with controls (mean at 6 months: 3.3 vs. 5.4; p<0.001 on scale scored with range 0-8; mean at 1 year: 21.0 vs. 24.5; p=0.0001 on scale scored with range 0-45).	Low
Global improvement	3 RCT (n=727)	All 3 trials report better global improvement for CBT vs. control (41 to -70% improved in CBT vs. 15-32% in controls).	Moderate

Key question Outcome	Study design Number of studies (n)*	Findings and direction of effect	Strength of evidence grade
Acclydine vs. placebo			
Improved overall function, decreased fatigue, and increased physical activity (actometer)	1 RCT (n=57)	No significant differences between intervention and placebo.	Insufficient
Improved quality of life, increased days spent at work/school, proportion working full-or part-time	No studies	No studies.	Insufficient
Acetyl-L-carnitine vs. propionyl-L-carnitine vs. combination			
Decreased fatigue	1 RCT (n=89)	Acetyl-L-carnitine had lower fatigue scores at 24 weeks, but propionyl-L-carnitine and the combination group improved more from baseline (p=0.004 and p<0.001, respectively).	Insufficient
Global improvement	1 RCT (n=89)	Significant improvement in propionyl-L-carnitine (63%, p<0.001) and acetyl-L-carnitine (59%, p<0.001) compared with the combination group (37%, p=0.084).	Insufficient
Improved overall function, quality of life, increased days spent at work/school, proportion working full- or part-time	No studies	No studies.	Insufficient
Pollen extract vs. placebo			
Decreased fatigue	1 RCT (n=22)	Significant improvement on fatigue scores in the pollen group compared with placebo at 3 months (-0.43 vs. -0.18, p<0.05).	Insufficient
Improved quality of life	1 RCT (n=22)	Significant improvement in quality of life scores in the pollen group compared with placebo at 3 months (-1.66 vs. -0.21; p<0.01).	Insufficient
Improved overall function, increased days spent at work/school, proportion working full- or part-time	No studies	No studies.	Insufficient
Low sugar/low yeast diet vs. healthy eating			
Decreased fatigue, improved quality of life	1 RCT (n=39)	No significant differences between interventions.	Insufficient
Improved overall function, increased days spent at work/school, proportion working full- or part-time	No studies	No studies.	Insufficient

Key question Outcome	Study design Number of studies (n)*	Findings and direction of effect	Strength of evidence grade
Distant healing vs. no treatment			
Improved overall function	1 RCT (n=409)	Significant improvement on functioning scores for those who were blinded to the treatment compared with those who were not blinded to the treatment (covariance analysis effect for blinded vs. unblinded treatment: -1.54 [SE 0.70] 95% CI -2.91 to -0.18). No other significant differences between intervention and no treatment.	Insufficient
Decreased fatigue, improved quality of life, increased days spent at work/school, proportion working full- or part-time	No studies	No studies.	Insufficient
Homeopathy vs. placebo			
Decreased fatigue	1 RCT (n=89)	Significantly better scores on MFI-20 for placebo group compared with intervention at 6 months (mean: 2.70 vs. 1.35, p=0.04).	Insufficient
Improved overall function	1 RCT (n=89)	No significant differences between intervention and placebo.	Insufficient
Improved quality of life, increased days spent at work/school, proportion working full-or part-time	No studies	No studies.	Insufficient
Melatonin vs. phototherapy			
Improved overall function and decreased fatigue	1 RCT crossover design (n=30)	No significant differences between interventions.	Insufficient
Improved quality of life, increased days spent at work/school, proportion working full-or part-time	No studies	No studies.	Insufficient
Home orthostatic training vs. sham home orthostatic training			
Improved overall function	1 RCT (n=36)	No significant differences between interventions.	Insufficient
Decreased fatigue, improved quality of life, increased days spent at work/school, proportion working full- or part-time	No studies	No studies.	Insufficient

Key question Outcome	Study design Number of studies (n)*	Findings and direction of effect	Strength of evidence grade
Qigong exercise vs. no qigong exercise			
Improved overall function	1 RCT (n=52)	Significantly better SF-12 physical functioning scores for qigong exercise compared with no exercise at 4 months (mean: 42.7 vs. 35.7, p=0.001).	Insufficient
Decreased fatigue	1 RCT (n=52)	Significantly better Chalder Fatigue Scale scores in exercise group compared with no exercise group at 4 months (mean total: 21.6 vs. 32.1, p<0.001; mean physical fatigue subscale: 12.9 vs. 20.3, p<0.001; mean mental fatigue subscale: 8.8 vs. 11.9, p=0.012).	Insufficient
Improved quality of life, increased days spent at work/school, proportion working full-or part-time	No studies	No studies.	Insufficient
GET vs. no treatment or flexibility/relaxation therapy or adaptive pacing			
Improved overall function	4 RCT (n=619)	Results from 3 studies that used the SF-36 physical functioning subscale were pooled, there was a significant effect for the intervention group to have better scores vs. control at followup: weighted mean difference 10.29 (95% CI 6.71 to 13.86).	Moderate
Decreased fatigue	4 RCT(n=619)	Significantly better Chalder Fatigue Scale scores reported for exercise groups compared with controls in 3 of the studies:Mean total: 13.91 vs. 24.41; p=0.02, physical fatigue scores: 7.91 vs. 14.27; p=0.02; and mental fatigue scores: 6.00 vs. 10.14; p=0.03 at 12 weeks; mean total: 20.5 vs. 27.4; p=0.004 at 12 weeks; and mean difference in change from baseline from adaptive pacing: -2.5; 95% CI -4.2 to -0.9; p=0.0059 and no treatment: -3.4; 95% CI -5.0 to -1.8; p=0.0001.1 study reported no differences between groups.	Low
Increased proportion working full- or part-time	1 RCT (n=59)	More in the exercise group were working at 1 year compared with control (66% vs. 39%; 95% CI 9% to 44%).	Insufficient
Decreased work impairment	1 RCT (n=475)	Significant improvement reported for exercise group on work and social adjustment scale compared with adaptive pacing and no treatment at 1 year (20.5 vs. 24.5 vs. 23.9; p=0.0004 and p<0.001, respectively).	Low

Key question Outcome	Study design Number of studies (n)*	Findings and direction of effect	Strength of evidence grade
Global improvement	3 RCT (n=583)	Significantly more improvement reported in exercise groups (31% and 54%) compared with controls (7%, p=0.05 and 24%, p=0.04). RR 1.54 (95% CI 1.26 to 1.89)	Moderate
Improved quality of life, increased days spent at work/school	No studies	No studies.	Insufficient
GET ± fluoxetine vs. fluoxetine ± placebo			
Improved overall function	1 RCT (n=136)	Significant improvement for exercise groups (either alone or combination) on functional work capacity at 26 weeks (mean change from baseline: 1.9; 95% CI 0.15 to 3.69; p=0.03) compared with other groups.	Low
Decreased fatigue	1 RCT (n=136)	Significantly more individuals in exercise groups (either alone or combination) did not meet the threshold of "caseness" for fatigue on Chalder Fatigue Scale (18% for both exercise groups and 6% for both other groups; p=0.025).	Low
Increased days spent at work/school and proportion working full- or part-time	No studies	No studies.	Insufficient
Face-to-face CBT vs. telephone CBT			
Clinical global improvement	1 RCT (n=65)	More individuals rated as much better or very much better in face-to-face group compared with telephone group (6 months: 60% vs. 40%; p=NR and 12 months: 57% vs. 55%; p=NR).	Insufficient
Improved overall function, decreased fatigue and work impairment	1 RCT (n=65)	No significant differences between interventions.	Insufficient
Quality of life, days spent at work/school, proportion working full-or part-time	No studies	No studies.	Insufficient
CBT + GET vs. usual care			
Improved overall function, and decreased fatigue	1 RCT (n=115)	No significant differences between intervention and control.	Insufficient
Improved quality of life, decreased work impairment, increased days spent at work/school, proportion working full-or part-time	No studies	No studies.	Insufficient

Key question Outcome	Study design Number of studies (n)*	Findings and direction of effect	Strength of evidence grade
KQ2b. What are the harms of therapeutic interventions for patients with ME/CFS and how do they vary by patient subgroups?			
Galantamine vs. placebo	1 RCT (n=434)	90% (389/434) reported harms; 23% (88/389) withdrew due to harms; 2% (8/389) in galantamine reported serious harms but none attributed to the study drug; no significant differences reported between groups.	Insufficient
Hydrocortisone vs. placebo	1 RCT (n=70)	More harms reported with hydrocortisone vs. placebo (suppression of adrenal glucocorticoid responsiveness: 12 vs. 0; p<0.001; increased appetite: 17 vs. 8; p=0.02; weight gain: 19 vs. 8; p=0.006; difficulty sleeping: 17 vs. 8; p=0.02); no other significant differences between groups.	Insufficient
Hydrocortisone + fludrocortisone vs. placebo	1 RCT (n=80)	1.3% (1/80) withdrew due to acne and weight gain, no serious harms reported; no other harms data reported.	Insufficient
Immunoglobulin G vs. placebo	1 RCT (n=28)	Significantly more with headaches in immunoglobulin G group vs. placebo (93% vs. 60%; p=0.03); 20% total harms overall; 1 in each group withdrew due to harms; 2 in immunoglobulin G and 3 in placebo developed serious harms.	Insufficient
Rintatolimod vs. placebo	2 RCT (n=324)	Flu-like syndrome, chills, vasodilatation, and dyspnea were more frequent in rintatolimod vs. placebo (p<0.05); no other differences between groups.	Insufficient
Valganciclovir vs. placebo	1 RCT (n=30)	No one withdrew due to harms, 1 in each group developed cancer, deemed unrelated; no other harms data reported.	Insufficient
Isoprinosine vs. placebo	1 RCT (n=15)	No one withdrew due to harms; no other harms data reported.	Insufficient
Fluoxetine vs. placebo	1 RCT (n=68)	More total withdrawals in the fluoxetine group compared with placebo.	Insufficient
CBT/counseling vs. no treatment or support or relaxation or adaptive pacing			
Withdrawals due to harms	1 RCT (n=47)	1 trial reported none withdrew due to harms.	Insufficient
Rates of harms	1 RCT (n=257)	1 trial reported no differences between groups for reported harms.	Insufficient
Total harms	1 RCT (n=471)	1 large trial reported fewer total harms in the CBT group (848) vs. adaptive pacing (949) and no treatment (977), but p=NR. The other study did not report harms by group, but deemed all unrelated to the intervention.	Low

Key question Outcome	Study design Number of studies (n)*	Findings and direction of effect	Strength of evidence grade
Serious harms	2 RCT (n=728)	1 large trial (n=471) reported fewer serious harms in the CBT group per 100 person-years (5.0; 95% CI 2.2 to 9.8) vs. adaptive pacing (10.1; 95% CI 5.8 to 16.3), but was similar to no treatment (4.4; 95% CI1.8 to 9.0). The other trial reported that no serious harms were reported.	Low
Acclydine vs. placebo	No studies	No studies.	Insufficient
Acetyl-L-carnitine vs. propionyl-L-carnitine vs. combination	1 RCT (n=89)	No differences reported between groups for withdrawals due to harms; no other harms data reported.	Insufficient
Pollen extract vs. placebo	No studies	No studies.	Insufficient
Low sugar/low yeast diet vs. healthy eating	No studies	No studies.	Insufficient
Distant healing vs. no treatment	No studies	No studies.	Insufficient
Homeopathy vs. placebo	No studies	No studies.	Insufficient
Melatonin vs. phototherapy	No studies	No studies.	Insufficient
Home orthostatic training vs. sham home orthostatic training	No studies	No studies.	Insufficient
Qigong exercise vs. no qigong exercise	1 RCT (n=52)	No harms were reported by either group, no other harms data provided.	Insufficient
GET vs. no treatment or flexibility/relaxation therapy or adaptive pacing			
Withdrawals due to harms	1 RCT (n=49)	1 trial reported 40% (10/25) of GET group refused to repeat the required fitness test due to feeling initial test was harmful and 1 person withdrew due to a calf injury.	Insufficient
Total harms	2 RCT (n=524)	1 trial reported similar harms in the GET group (992) vs. adaptive pacing (949) and no treatment (977), but p=NR. The other trial reported 2% (1/49) experienced a harm.	Low
Serious harms	1 RCT (n=475)	1 large trial reported similar serious harms in GET group per 100 person-years (10.6; 95% CI 6.2 to 17.0) vs. adaptive pacing (10.1; 95% CI 5.8 to 16.3) but fewer in no treatment (4.4; 95% CI 1.8 to 9.0).	Low
GET vs. fluoxetine vs. combination or placebo	1 RCT (n=136)	11 withdrawals due to medication side effects 13% in fluoxetine group vs. 3% in placebo group; no other harms data reported in study.	Insufficient
Face-to-face CBT vs. telephone CBT	No studies	No studies.	Insufficient
CBT + GET vs. usual care	No studies	No studies.	Insufficient

Key question Outcome	Study design Number of studies (n)*	Findings and direction of effect	Strength of evidence grade
KQ2c. What are the characteristics of responders and non-responders to interventions?			
CBT vs. no treatment			
Baseline differences	1 RCT (n=27)	Significant differences between those who responded to CBT and those who did not on baseline measures of functional impairment on SIP-8 (mean: 1,330 vs. 1,985; p=0.031), daily observed fatigue (mean on scale 0-16: 7.4 vs. 9.7; p=0.023), and daily observed pain (mean on scale 0-16: 4.5 vs. 7.8; p=0.026); but not for hours worked per week (mean: 10.9 vs. 2.6; p=0.062).	Insufficient

* Sample size includes only those analyzed

CBT= Cognitive Behavioral Therapy; CFS = chronic fatigue syndrome; CI= Confidence Interval; CIS= Checklist of Individual Strength; FSS= Fatigue Severity Scale; GET= graded exercise therapy; KPS = Karnofsky Performance Score; KQ= key question; ME = Myalgic encephalomyelitis; MFI-20= Multidimensional Fatigue Inventory; n= sample size; NR= not reported; QLI= Quality of Life Index; RCT = randomized controlled trial; SE= standard error; SF-12= Short Form 12-item Health Survey; SF-36= 36-item Short Form Survey; SIP-8 = Sickness Impact Profile 8-items; SOFA-CFS= Schedule of Fatigue and Angina for CFS scale; vs.= versus.

Findings in Relationship to What is Already Known

The lack of a clear etiology for ME/CFS, the multisystem involvement of the syndrome, and its overlap with other chronic conditions contribute to the difficulty in diagnosing ME/CFS and the risk of misdiagnosing a patient with an overlapping condition or incorrectly labeling a patient with ME/CFS. Much research in this field focuses on discovering etiologies rather than testing diagnostic strategies in patients. Articles that attempted to define an etiology on the basis of a biochemical marker or a particular physiologic test were not included in this review because the intent of these was to identify an etiology rather than understand how the specific test could distinguish patients that would respond to treatment. In addition to biomarker studies (cell function, immunologic, virologic/bacteriologic, hormonal, etc.), studies identified subgroups on the basis of exercise testing,[110,111] cerebral blood flow as measured by arterial spin labeling,[112] gait kinetics,[113] impaired blood pressure variability/hemodynamic instability,[114,115] bioenergetics (capacity to recover from acidosis),[116] and many others. These studies did not report diagnostic testing outcomes such as ROC/AUC, sensitivity, specificity, or concordance and were therefore not included in the diagnostic testing section of this report. The studies on serum parameters and cardiopulmonary function/recovery that did meet the inclusion criteria were not adequately tested in a broad spectrum of patients to determine utility for distinguishing patients with ME/CFS compared with other patients with chronic and disabling conditions.

One of the primary limitations in the literature about diagnostic tests, was that very few studies included a validation cohort. Instead, these studies primarily evaluated a diagnostic test in a single initial population (a derivation cohort). Derivation studies are a necessary first step when attempting to achieve a valid diagnostic test, but they also have inherent methodological problems. They often involve the use of cases and controls – two very distinct populations – in order to determine whether the test can distinguish between those two groups. If the test is capable of distinguishing between two distinct groups, then further testing should use populations that are more closely related (i.e., they have overlap in terms of symptoms), in order to more rigorously test the diagnostic capability of a particular test. As such, the more rigorous diagnostic testing studies will include a population for whom the clinician is likely to face diagnostic uncertainty, and then test how well the test performs in classifying that population accurately. The studies identified for evaluation of diagnostic tests for ME/CFS fell into three main categories. The first are those that evaluated a diagnostic test or a scale against a chosen reference standard. In this case, the reference standard was typically one or more of several case definitions that have been published (CDC Holmes, 1988 and CDC Fukuda, 1994, Canadian ME/CFS definition, International Consensus Criteria for ME, etc.). A second group of studies evaluated how those case definitions compare with each other, and whether they identify the same or different populations. While this was not a distinct key question, it was felt to shed light on the evolving definition of ME/CFS and the difficulty with identifying a universally acceptable reference standard. The third group of studies presented here are those that address harms of diagnosis.

ME/CFS is a condition that lacks a universally accepted diagnostic (gold) standard, a criterium that defines the condition. The lack of gold standard poses significant challenges for evaluation of diagnostic tests, and yet this is a situation that arises commonly with conditions that are syndromes. A syndrome is a "combination of symptoms and signs which have been observed to occur together so frequently and to be so distinctive that they constitute a recognizable clinical picture.[117] That is to say that the combination of findings is unusual so as

not to be thought of as coincidence. In such situations, the traditional evaluation of a diagnostic test is more challenging. The ME/CFS literature is beginning to test diagnostic strategies but as yet has not presented data that would sufficiently differentiate the diagnosis of ME/CFS from other similar conditions in a population of patients with substantial diagnostic uncertainty. For example, a proposed test might sufficiently distinguish a patient with ME/CFS from one without, but may not be able to distinguish between a patient with ME/CFS and one with depression or rheumatoid arthritis – conditions that a clinician might be considering simultaneously and attempting to rule out in a patient who presents with fatigue. There were no studies that quantitatively compared the diagnostic concordance of two case definitions. Several studies attempted to demonstrate that ME, ME/CFS, and CFS case definitions identify different groups of people. Studies did this by identifying people who met one criteria but not the other.[4,5,53,54,58] Using this approach, it appears that the ME and ME/CFS case definitions select a population with more impairment, lower functioning, and higher symptom reporting compared with CFS alone. Other studies compared subjects who met a definition of CFS with subjects who had other disease states and/or those comprising a healthy control population.[52,55,56] As expected, these studies demonstrated CFS subjects have lower functioning and higher symptom burdens than people from the general population.

One systematic review compared case definitions for ME/CFS using a slightly different approach. This review summarized how the prevalence of ME/CFS in a population and the symptom burden for patients varies when using different case definitions.[118] This study attempted to bring some consistency to case definitions for ME/CFS in the absence of a reference standard. Their inclusion criteria were broader than those for this report and similarly, they found that the validation studies were weak and heterogeneous. This group called for the community of ME/CFS researchers to prioritize research on treatments using existing case definitions (of which, they felt the CDC Fukuda, 1994 criteria had the most studies on validation and comparison with other measures, and was thought to be the most appropriate for clinical practice), rather than development of additional new case definitions.[118]

Patients with ME/CFS report feeling stigmatized by their diagnosis in terms of financial stability, work opportunities, perceived judgments on their character, social isolation, and interactions with health care providers. Compounding this is the substantial burden of misdiagnosis among this patient population. Two studies objectively identified prejudice and stereotypes towards patients with ME/CFS from members of the medical community; medical trainees and mental health practitioners make judgments about a patient's condition based on the name it carries (ME, CFS, or other) and what treatment is being given. While these studies were descriptive and based on survey data, the results suggest valid concerns about the harm of labeling patients with a diagnosis of ME/CFS. These harms may reflect the chronic and disabling nature of this disease, combined with a lack of understanding about the diagnosis among the medical community and uncertainty about the etiology of ME/CFS. One commentary suggested that the harm is associated with the implications of a label rather than the label itself, and that it is "acceptable and often beneficial to make diagnoses such as CFS, provided that this is the beginning and not the end, of the therapeutic encounter."[119]

Determining the efficacy of medication and CAM interventions to treat ME/CFS was limited because most were only evaluated in single studies which had significant methodological limitations including small sample sizes. Additionally, outcomes were assessed using different methods and different scales. Trials were generally designed as pilot studies at single centers and were small, with some enrolling fewer than 20 subjects in an arm. Some medication trials were

primarily intended to measure intermediate outcomes, such as natural killer cell-mediated cytotoxicity,[74] and most were underpowered for the health outcomes relevant to this systematic review. While several fatigue and function outcomes were based on validated scales and measures, others were not, and the clinical significance of changes in scores over time are not clear.

Although placebo-controlled trials of immune modulating and antiviral medications suggested potential improvement in fatigue and functioning, some findings were of borderline statistical significance and other outcomes did not differ between groups. The rationale for treating patients with medications that have antiviral or immunomodulatory properties is based on the association of ME/CFS with viruses and immunological abnormalities that may underlie or promote its pathogenesis.[13,120-122] Although small trials of acyclovir,[123] immunoglobulin G,[70,124] and Isoprinosine[74] indicated no statistically significant differences between treatment and placebo groups for measures of fatigue, quality of life, or function, two trials of intravenous rintatolimod[72,73] and a trial of oral valganciclovir[71] suggested improvement. These trials differed from the earlier trials by using newer medications and applying selective inclusion criteria for participants that targeted patient subgroups based on clinical history of a likely viral onset of ME/CFS and high antibody titers[71] or severe disability.[72,73] In addition, the two rintatolimod trials were much larger than the others providing stronger statistical power to detect differences between groups. However, these studies are not definitive and are limited by inconsistencies in methods and findings, small numbers, methodological shortcomings, and lack of long term followup. Trials of galantamine, hydrocortisone, and immunoglobulin G indicated no significant improvement compared with placebo. Harms related to medications that were statistically significantly higher for the treatment versus placebo groups included suppression of adrenal glucocorticoid responsiveness, increased appetite, weight gain, and difficulty sleeping with hydrocortisone; flu-like syndrome, chills, vasodilatation, dyspnea, and dry skin with rintatolimod; and headaches with immunoglobulin G.

Consistent with other systematic reviews, both CBT and GET were found to improve symptoms, primarily based on fatigue and function outcomes, whereas evidence on other non-pharmacological interventions was inconclusive.[125-128] Results need to be interpreted with caution given that studies often used multiple methods of evaluating their outcomes and several had mixed results on the same measure when comparing different tools. Although some of the studies attempted to measure adherence, inherent inaccuracies exist with self-reporting particularly when it applies to home exercise programs.

Harms were not well reported throughout all of the non-pharmacological and CAM interventions. Although a previous systematic review suggested that the placebo response in the treatment of CFS is lower (19.5%) than would be expected (30-50%),[129] the results of the distant healing trial suggest that expectation theory, a patient's expectation and belief of a positive or negative result, may influence the outcome in treatment trials of ME/CFS patients.[103] When reported, the harms associated with exercise included total harms, serious harms, or withdrawal due to harms but the specific harms were not delineated.[75,84,98,108] In the combination trials, the greatest number of harms were in the GET arm of one trial,[98] lowest adherence was in the exercise arm in another trial,[84] and several trials had greatest withdrawal due to harms in the exercise arms.[75,106,108]. The higher rates of refusing to repeat physiological testing implies significant harm in at least some of the participants.[108] Several previous studies have found worsening effects with exercise and a survey sponsored by the ME Association found that patients believed that GET made more people worse compared with other treatments.[130,131] One

study comparing CBT with cognitive therapy, anaerobic exercise, or relaxation found that those patients who remained within their energy envelope (avoided overexertion and under exertion by exerting a comfortable range of energy) had a significant improvement in mean fatigue and functioning scores regardless of treatment arm.[82]

One gap in the body of the evidence is the lack of subgroup analysis based on factors such as clinical features at baseline (extent of PEM, autonomic dysfunction, neurocognitive impairment, etc.), severity of disease, duration of disease, and patient demographics. Effectiveness and/or harms may differ between patient subgroups and given the small sample size of most of the trials, combining all patients may have lessened the effect size. A recent systematic review that compared different case definitions agreed that patients should be classified according to their severity and symptom patterns in order to optimally guide therapy and predict prognosis.[118]

Applicability

The applicability of our findings to real-world clinical settings is supported by several features of the body of literature we reviewed. First, we included all recognized case definitions of ME/CFS in order to allow a broad representation of patients. Studies were conducted primarily in the United States or Western Europe and patients had a female predominance which is consistent with clinical practice. Duration of symptoms, while not consistently reported, was broadly represented across studies. The interventions and comparators represented most of the therapeutic modalities commonly used in clinical practice.

There are however several features of this body of evidence that limit its generalizability to the broader population of patients with ME/CFS, including factors surrounding the diagnosis itself. Given that the condition is a syndrome with a constellation of symptoms and lacking a gold standard for diagnostic comparison, it is at inherent risk of bias by the opinion of experts. For example, experts have identified critical features of the condition including PEM, however current methods of testing, comparing, and monitoring this symptom are lacking. Many of the diagnostic studies have also been conducted in a referral based environment and lacking a spectrum of patients, some with and some without the disease. Patients tended to be white middle-aged women and it is unknown if the results are generalizable to other demographic populations. Most of the intervention trials have small sample size, few treatments being evaluated in greater than one study, and rarely reporting baseline function or severity of illness. Patients from specialty clinics may represent a more severe form of the condition. Patients from rural centers or lacking insurance or finances may not have access to specialty clinics or clinical trials. Additionally, although most trials included patients based on the CDC (Fukuda, 1994) case definition, some included other diagnostic criteria. We elected to include trials using any pre-defined case definition but recognize that some of the earlier criteria, in particular the Oxford (Sharpe, 1991) criteria, could include patients with 6 months of unexplained fatigue and no other features of ME/CFS. This has the potential of inappropriately including patients that would not otherwise be diagnosed with ME/CFS and may provide misleading results. Finally, given the heterogeneity in the outcomes evaluated and the methods of measuring the outcomes of interest, quantitative meta-analysis could not be performed, and comparison between studies is limited.

Implications for Clinical and Policy Decisionmaking

The limitations in applicability, as well as the limitations of the evidence base, make it difficult to draw firm conclusions with implications for clinical practice. Studies surrounding aspects of diagnosis suggest that different case definitions introduce variability in the

characteristics of the population identified as having ME/CFS and that some of the case definitions will be more inclusive (including patients with overlapping conditions) whereas others may be more specific but identify more severe forms of the condition albeit a smaller population. No one tool has been adequately studied or identified to clearly discriminate between patients particularly when there is diagnostic uncertainty.

Most of the evidence available surrounding treatment is insufficient to draw conclusions. Because of limitations in the evidence base, we did not have high confidence in any of the findings from this review, and only had moderate confidence in the benefit of CBT (fatigue and global improvement) and GET (function and global improvement). In clinical practice, treatment of ME/CFS often involves multiple concurrent therapies but we found few trials that compared one intervention to another or that compared a combination of concurrent therapies with another. The trial on valganciclovir, an antiviral medication pre-selected patients with an inciting febrile event with lymphadenopathy suggests improvement in fatigue in this population of ME/CFS while the trials on immune modulators, that included patients who were severely disabled, found some improvement in exercise capacity. Both CBT and GET showed improvement in most outcomes but the combination of CBT and GET has not been adequately studied (one trial) to determine if this is more effective than a single intervention. Subgroup analysis in the GET trials would help in identifying if there are specific patients who might have greater benefit or experience greater harm. Other interventions have not been adequately studied (single small trials, heterogeneity in outcomes measured) to guide clinical and policy decision making.

Across all intervention trials, heterogeneity in the population samples (different case definitions used for inclusion), outcomes evaluated, and tools used to measure these outcomes, limited the ability to synthesize data. Acceptance of a single case definition and development of a core outcome set would aide in better studying the interventions to allow for more meaningful guidance for clinicians, policy makers, and patients.

Limitations of the Evidence Base

What are the limitations of the evidence?

The most important potential limitation of our review is that important studies whose findings might influence clinical and policy decision making may not have been identified. We conducted a comprehensive, broadly inclusive search that produced 5,902 study titles and abstracts. Although we excluded non-English language studies and studies published before 1988, we do not believe that important studies of therapies used in current practice were missed; the general consistency of our findings with other systematic reviews, provides some assurance that our review was not biased by our selection criteria. Our review focused on diagnostic methods that provided data on a test's utility in identifying patients with ME/CFS (ROC/AUC, sensitivity, specificity, concordance). Other testing strategies were not reviewed and may provide further insight methods of identifying patients with ME/CFS. To evaluate the effectiveness and harms of interventions, we elected to include studies of 12 weeks or longer duration due to the cyclical nature of the condition. Notably, often antiviral or antibiotic medications are traditionally prescribed for a shorter duration and would not have been included in our report. To account for this, we performed a concurrent search for antiviral therapies in the treatment of ME/CFS and only found one additional trial of shorter duration that did not change our results.[123] We considered outcomes of overall improvement, fatigue, function, quality of life, and employment which we considered clinically significant and conducive to the systematic review methodology. Given the breadth of symptoms in ME/CFS, we a priori elected to not review symptom related outcomes except for fatigue. Some interventions may have revealed benefit for other characteristics of ME/CFS and this review would not have identified these outcomes.

There may have been biased reporting of results in the literature such that only selected studies were published and retrievable, and that published studies may have been affected by conflicts of interest, outcome reporting bias or analysis reporting bias. Reporting bias and conflicts of interest are concerns with any systematic review. We were not able to conduct quantitative analyses to evaluate the possibility of publication bias for our findings because of the heterogeneity across studies in our review, and in many cases the lack of key information needed to perform quantitative syntheses, generally precluded meaningful comparison of effect sizes. Mitigating against the likelihood of publication bias in our review, however, is the fact that the majority of studies in our review were small (most <100 patients, many <50) and most reported no significant effect of the intervention. Publication bias typically results in selective publication of larger studies and/or those with positive findings, and studies biased by conflict of interests would also be more likely to report positive findings. We also conducted gray literature searches to look for unpublished data and did not find evidence of unreported studies. We did not have access to study protocols, such that our assessment of outcome and analysis reporting bias were limited. However, the limited and vague reporting of harms in many studies may suggest outcome reporting bias for these outcomes.

The main limitation of the evidence base in our review was poor study quality. Most trials did not specify randomization method, did not conceal allocation, and did not mask outcomes assessment. Most studies were small and many were underpowered to detect significant differences. Studies were also highly variable in terms of methods used to measure outcomes. limiting our ability to combine or compare results across studies.

Future Research

What are the future research needs for definition, diagnosis, and treatment of ME/CFS?

Future studies evaluating the diagnostic capability of instruments for the identification of ME/CFS should include populations that include a broad range of people with a full complement of conditions that require clinical distinction from ME/CFS, such as fibromyalgia and depression. Thus, the ideal diagnostic test for ME/CFS would adequately distinguish between ME/CFS and these conditions. Additionally, studies should report statistics on how well a particular measure distinguishes a group with ME/CFS from a group that does not meet these criteria – using concordance and the net reclassification index. For physiological and metabolic testing, selection of a broader spectrum of patients as a comparative group is needed rather than healthy controls.

To inform clinical practice and policy, it would be ideal if future intervention studies consistently used an agreed upon single case definition to reduce variability in the patient samples. Trials should use multicomponent treatments, larger sample sizes with calculated power calculations and more rigorous adherence to methodological standards for clinical trials or observational studies. Given the cyclical nature of the condition, followup periods greater than one year would be optimal to determine the effectiveness over time. Given the plethora of outcome measures development of a set of core outcomes including more patient-centered outcomes such as quality of life, employment, and time spent supineversus active, would help guide research and facilitate future data syntheses. Reporting of information about cointerventions, the timing of studied interventions in relation to other interventions, and adherence to interventions would improve the applicability of study findings. Similarly, stratification of findings by patient characteristics (e.g., baseline severity, comorbidities, demographics) would help determine the applicability of different interventions for specific patients and situations. It is particularly important for future studies to report findings according to the cardinal features of ME/CFS such as PEM, neurocognitive status, and autonomic function

as treatment choices may differ for subsets of the population. Clearly reporting harms particularly surrounding exercise testing and treatment for specific subgroups may help identify patients more negatively affected by these interventions.

Conclusions

Multiple case definitions for ME/CFS exist with those that require symptoms of PEM, neurological impairment, and autonomic dysfunction representing a more severe form of the condition. No current diagnostic tool or method has been adequately tested to identify patients when diagnostic uncertainty exists. Although CBT and GET have shown benefit in some measures of fatigue, function, and global improvement, most other interventions have insufficient evidence to direct clinical practice. GET appears to be associated with harms in some patients whereas the negative effects of being given a diagnosis of ME/CFS appear to be more universal.

References

1. Carruthers BM, Jain AK, de Meirleir KL, et al. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: Clinical Working Case Definition, Diagnostic and Treatment Protocols. J Chronic Fatigue Syndr. 2003;11(1): 7-115.

2. Carruthers BM, van de Sande MI, De Meirleir KL, et al. Myalgic encephalomyelitis: International Consensus Criteria. J Intern Med. 2011;270(4): 327-38. PMID: 21777306.

3. Fukuda K, Straus SE, Hickie I, et al. The chronic fatigue syndrome: a comprehensive approach to its definition and study. International Chronic Fatigue Syndrome Study Group. Ann Intern Med. 1994;121(12): 953-9. PMID: 7978722.

4. Jason LA, Brown A, Evans M, et al. Contrasting chronic Fatigue syndrome versus myalgic encephalomyelitis/chronic fatigue syndrome. Fatigue. 2013;1(3)PMID: 23914329.

5. Jason LA, Brown A, Clyne E, et al. Contrasting case definitions for chronic fatigue syndrome, Myalgic Encephalomyelitis/chronic fatigue syndrome and myalgic encephalomyelitis. Eval Health Prof. 2012;35(3): 280-304. PMID: 22158691.

6. Holmes GP, Kaplan JE, Gantz NM, et al. Chronic fatigue syndrome: a working case definition. Ann Intern Med. 1988;108(3): 387-9. PMID: 2829679.

7. Ramsay M. Myalgic Encephalomyelitis and Postviral Fatigue States: The saga of Royal Free disease. 1st ed. London: Gower Medical Publishing; 1986.

8. Jason LA, Taylor RR, Plioplys S, et al. Evaluating attributions for an illness based upon the name: chronic fatigue syndrome, myalgic encephalopathy and Florence Nightingale disease. Am J Community Psychol. 2002;30(1): 133-48. PMID: 11928774.

9. Komaroff AL, Cho TA. Role of infection and neurologic dysfunction in chronic fatigue syndrome. Semin Neurol. 2011;31(3): 325-37. PMID: 21964849.

10. Morris G, Maes M. A neuro-immune model of Myalgic Encephalomyelitis/Chronic fatigue syndrome. Metab Brain Dis. 2013;28(4): 523-40. PMID: 22718491.

11. Hempel S, Chambers D, Bagnall AM, et al. Risk factors for chronic fatigue syndrome/myalgic encephalomyelitis: a systematic scoping review of multiple predictor studies. Psychol Med. 2008;38(7): 915-26. PMID: 17892624.

12. Chew-Graham C, Dowrick C, Wearden A, et al. Making the diagnosis of Chronic Fatigue Syndrome/Myalgic Encephalitis in primary care: a qualitative study. BMC Fam Pract. 2010;11: 16. PMID: 20178588.

13. Afari N, Buchwald D. Chronic fatigue syndrome: a review. Am J Psychiatry. 2003;160(2): 221-36. PMID: 12562565.

14. Anderson G, Maes M, Berk M. Biological underpinnings of the commonalities in depression, somatization, and Chronic Fatigue Syndrome. Med Hypotheses. 2012;78(6): 752-6. PMID: 22445460.

15. Reyes M, Nisenbaum R, Hoaglin DC, et al. Prevalence and incidence of chronic fatigue syndrome in Wichita, Kansas. Arch Intern Med. 2003;163(13): 1530-6. PMID: 12860574.

16. Reeves WC, Jones JF, Maloney E, et al. Prevalence of chronic fatigue syndrome in metropolitan, urban, and rural Georgia. Popul Health Metr. 2007;5: 5. PMID: 17559660.

17. Johnston S, Brenu EW, Staines D, et al. The prevalence of chronic fatigue syndrome/ myalgic encephalomyelitis: a meta-analysis. Clin Epidemiol. 2013;5: 105-10. PMID: 23576883.

18. Rangel L, Garralda ME, Levin M, et al. The course of severe chronic fatigue syndrome in childhood. J R Soc Med. 2000;93(3): 129-34. PMID: 10741312.

19. Crawley EM, Emond AM, Sterne JA. Unidentified Chronic Fatigue Syndrome/myalgic encephalomyelitis (CFS/ME) is a major cause of school absence: surveillance outcomes from school-based clinics. BMJ Open. 2011;1(2): e000252. PMID: 22155938.

20. Jason LA, Porter N, Hunnell J, et al. A natural history study of chronic fatigue syndrome. Rehabil Psychol. 2011;56(1): 32-42. PMID: 21401284.

21. Cairns R, Hotopf M. A systematic review describing the prognosis of chronic fatigue syndrome. Occup Med. 2005;55(1): 20-31. PMID: 15699087.

22. Center for Drug Evaluation and Research. The Voice of the Patient: A series of reports from the U.S. Food and Drug Administration's (FDA's) Patient-Focused Drug Development Initiative. Chronic Fatigue Syndrome and Myalgic Encephalomyelitis: U.S. Food and Drug Administration. 2013. Availible at: http://www.fda.gov/downloads/ForIndustry/UserFees/PrescriptionDrugUserFee/UCM368806.pdf. Accessed March 22, 2014.

23. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Publication No. 10(13)-EHC063-EF. Rockville (MD) :Agency for Healthcare Research and Quality. January 2014. Availible at: www.effectivehealthcare.ahrq.gov.

24. Smith MEB, Nelson N, Haney E, et al. Diagnosis and treatment of myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). PROSPERO 2014:CRD42014009779 2014. Availible at: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42014009779. Accessed July 3, 2014.

25. Hosmer Jr DW, Lemeshow S. Applied logistic regression: John Wiley & Sons; 2004.

26. Nisenbaum R, Jones JF, Unger ER, et al. A population-based study of the clinical course of chronic fatigue syndrome. Health Qual Life Outcomes. 2003;1: 49. PMID: 14613572.

27. Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. BMJ. 1994;309(6948): 188. PMID: 8044101.

28. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem. 1993;39(4): 561-77. PMID: 8472349.

29. Methods Guide for Medical Test Reviews. AHRQ Publication No. 12-EC017. Rockville, MD: Agency for Healthcare Research and Quality; June 2012. www.effectivehealthcare.ahrq.gov/reports/final.cfm. Also published as a special supplement to the Journal of General Internal Medicine, July 2012. PMID: 22834019.

30. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. Ann Intern Med. 2011;155(8): 529-36. PMID: 22007046.

31. Harris RP, Helfand M, Woolf SH, et al. Current methods of the US Preventive Services Task Force: a review of the process. Am J Prev Med. 2001;20(3 Suppl): 21-35. PMID: 11306229.

32. Atkins D, Chang SM, Gartlehner G, et al. Assessing applicability when comparing medical interventions: AHRQ and the Effective Health Care Program. J Clin Epidemiol. 2011;64(11): 1198-207. PMID: 21463926.

33. Fu R, Gartlehner G, Grant M, et al. Conducting quantitative synthesis when comparing medical interventions: AHRQ and the Effective Health Care Program. J Clin Epidemiol. 2011;64(11): 1187-97. PMID: 21477993.

34. Sutton A, Abrams K, Jones DR, et al. Methods for Meta-Analysis in Medical Research. Chichester, UK: John Wiley &Sons, Inc; 2000.

35. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21(11): 1539-58. PMID: 12111919.

36. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. BMJ. 2003;327(7414): 557-60. PMID: 12958120.

37. Tiev KP, Demettre E, Ercolano P, et al. RNase L levels in peripheral blood mononuclear cells: 37-kilodalton/83-kilodalton isoform ratio is a potential test for chronic fatigue syndrome. Clin Diagn Lab Immunol. 2003;10(2): 315-6. PMID: 12626460.

38. Linder R, Dinser R, Wagner M, et al. Generation of classification criteria for chronic fatigue syndrome using an artificial neural network and traditional criteria set. In Vivo. 2002;16(1): 37-43. PMID: 11980359.

39. Hadzi-Pavlovic D, Hickie IB, Wilson AJ, et al. Screening for prolonged fatigue syndromes: validation of the SOFA scale. Soc Psychiatry Psychiatr Epidemiol. 2000;35(10): 471-9. PMID: 11127722.

40. Jason LA, Evans M, Brown A, et al. Sensitivity and specificity of the CDC empirical chronic fatigue syndrome case definition. Psychology. 2010;1(1): 9-16. PMID: 23685416.

41. Jason L, Brown M, Evans M, et al. Measuring substantial reductions in functioning in patients with chronic fatigue syndrome. Disabil Rehabil. 2011;33(7): 589-98. PMID: 20617920.

42. Gaab J, Engert V, Heitz V, et al. Associations between neuroendocrine responses to the Insulin Tolerance Test and patient characteristics in chronic fatigue syndrome. J Psychosom Res. 2004;56(4): 419-24. PMID: 15094026.

43. Gaab J, Huster D, Peisen R, et al. Low-dose dexamethasone suppression test in chronic fatigue syndrome and health. Psychosom Med. 2002;64(2): 311-8. PMID: 11914448.

44. Gaab J, Rohleder N, Heitz V, et al. Stress-induced changes in LPS-induced pro-inflammatory cytokine production in chronic fatigue syndrome. Psychoneuroendocrinology. 2005;30(2): 188-98. PMID: 15471616.

45. Davenport TE, Stevens SR, Baroni K, et al. Reliability and validity of Short Form 36 Version 2 to measure health perceptions in a sub-group of individuals with fatigue. Disabil Rehabil. 2011;33(25-26): 2596-604. PMID: 21682669.

46. Davenport TE, Stevens SR, Baroni K, et al. Diagnostic accuracy of symptoms characterising chronic fatigue syndrome. Disabil Rehabil. 2011;33(19-20): 1768-75. PMID: 21208154.

47. Brown AA, Evans MA, Jason LA. Examining the energy envelope and associated symptom patterns in chronic fatigue syndrome: Does coping matter? Chronic Illn. 2013;9(4): 302-11. PMID: 23585632.

48. Reeves WC, Wagner D, Nisenbaum R, et al. Chronic fatigue syndrome--a clinically empirical approach to its definition and study. BMC Med. 2005;3: 19. PMID: 16356178.

49. Sharpe MC, Archard LC, Banatvala JE, et al. A report--chronic fatigue syndrome: guidelines for research. J R Soc Med. 1991;84(2): 118-21. PMID: 1999813.

50. Dowsett E, Goudsmit E, Macintyre A, et al. Report from The National Task Force on Chronic Fatigue Syndrome (CFS), Post Viral Fatigue Syndrome (PVFS), Myalgic Encephalomyelitis (ME). Westcare. 1994

51. Jason L, Evans M, Porter N, et al. The Development of a Revised Canadian Myalgic Encephalomyelitis Chronic Fatigue Syndrome Case Definition. Am J Biochem Biotechnol. 2010;6(2)

52. Aslakson E, Vollmer-Conna U, White PD. The validity of an empirical delineation of heterogeneity in chronic unexplained fatigue. Pharmacogenomics. 2006;7(3): 365-73. PMID: 16610947.

53. Brown AA, Jason LA, Evans MA, et al. Contrasting case definitions: The ME International Consensus Criteria vs. the Fukuda et al. CFS criteria. N Am J Psychol. 2013;15(1): 103-20.

54. Jason LA, Torres-Harding SR, Taylor RR, et al. A comparison of the 1988 and 1994 diagnostic criteria for chronic fatigue syndrome. J Clin Psychol Med Settings. 2001;8(4): 337-43.

55. Katon WJ, Buchwald DS, Simon GE, et al. Psychiatric illness in patients with chronic fatigue and those with rheumatoid arthritis. J Gen Intern Med. 1991;6(4): 277-85. PMID: 1890495.

56. Komaroff AL, Fagioli LR, Doolittle TH, et al. Health status in patients with chronic fatigue syndrome and in general population and disease comparison groups. Am J Med. 1996;101(3): 281-90. PMID: 8873490.

57. Lewis I, Pairman J, Spickett G, et al. Is chronic fatigue syndrome in older patients a different disease? -- a clinical cohort study. Eur J Clin Invest. 2013;43(3): 302-8. PMID: 23397955.

58. Van Hoof E, De Meirleir K. Chronic Fatigue Syndrome and Myalgic Encephalomyletis: Are Both Conditions on the Same Continuum? N Am J Psychol. 2005;7(2): 189-204.

59. Jason LA, Taylor RR. Measuring attributions about chronic fatigue syndrome. J Chronic Fatigue Syndr. 2001;8(3-4): 31-40.

60. Asbring P, Narvanen A-L. Women's experiences of stigma in relation to chronic fatigue syndrome and fibromyalgia. Qual Health Res. 2002;12(2): 148-60. PMID: 11837367.

61. Assefi NP, Coy TV, Uslan D, et al. Financial, occupational, and personal consequences of disability in patients with chronic fatigue syndrome and fibromyalgia compared to other fatiguing conditions. J Rheumatol. 2003;30(4): 804-8. PMID: 12672203.

62. Deale A, Wessely S. Diagnosis of psychiatric disorder in clinical evaluation of chronic fatigue syndrome. J R Soc Med. 2000;93(6): 310-2. PMID: 10911826.

63. Dickson A, Knussen C, Flowers P. Stigma and the delegitimation experience: An interpretative phenomenological analysis of people living with chronic fatigue syndrome. Psychol Health. 2007;22(7): 851-67.

64. Green J, Romei J, Natelson BH. Stigma and chronic fatigue syndrome. J Chronic Fatigue Syndr. 1999;5(2): 63-95.

65. Guise J, McVittie C, McKinlay A. A discourse analytic study of ME/CFS (Chronic Fatigue Syndrome) sufferers' experiences of interactions with doctors. J Health Psychol. 2010;15(3): 426-35. PMID: 20348363.

66. Jason LA, Taylor RR, Stepanek Z, et al. Attitudes regarding chronic fatigue syndrome: The importance of a name. J Health Psychol. 2001;6(1): 61-71. PMID: 22049238.

67. Blacker CVR, Greenwood DT, Wesnes KA, et al. Effect of galantamine hydrobromide in chronic fatigue syndrome: a randomized controlled trial. JAMA. 2004;292(10): 1195-204. PMID: 15353532.

68. McKenzie R, O'Fallon A, Dale J, et al. Low-dose hydrocortisone for treatment of chronic fatigue syndrome: a randomized controlled trial. JAMA. 1998;280(12): 1061-6. PMID: 9757853.

69. Blockmans D, Persoons P, Van Houdenhove B, et al. Combination therapy with hydrocortisone and fludrocortisone does not improve symptoms in chronic fatigue syndrome: a randomized, placebo-controlled, double-blind, crossover study. Am J Med. 2003;114(9): 736-41. PMID: 12829200.

70. Peterson PK, Shepard J, Macres M, et al. A controlled trial of intravenous immunoglobulin G in chronic fatigue syndrome. Am J Med. 1990;89(5): 554-60. PMID: 2239975.

71. Montoya JG, Kogelnik AM, Bhangoo M, et al. Randomized clinical trial to evaluate the efficacy and safety of valganciclovir in a subset of patients with chronic fatigue syndrome. J Med Virol. 2013;85(12): 2101-9. PMID: 23959519.

72. Strayer DR, Carter WA, Brodsky I, et al. A controlled clinical trial with a specifically configured RNA drug, poly(I) midline dot poly(C12U), in chronic fatigue syndrome. Clin Infect Dis. 1994;18(SUPPL. 1): S88-S95. PMID: 8148460.

73. Strayer DR, Carter WA, Stouch BC, et al. A double-blind, placebo-controlled, randomized, clinical trial of the TLR-3 agonist rintatolimod in severe cases of chronic fatigue syndrome. PLoS ONE. 2012;7(3): e31334. PMID: 22431963.

74. Diaz-Mitoma F, Turgonyi E, Kumar A, et al. Clinical improvement in chronic fatigue syndrome is associated with enhanced natural killer cell-mediated cytotoxicity: The results of a pilot study with Isoprinosine. J Chronic Fatigue Syndr. 2003;11(2): 71-93.

75. Wearden AJ, Morriss RK, Mullis R, et al. Randomised, double-blind, placebo-controlled treatment trial of fluoxetine and graded exercise for chronic fatigue syndrome.[Erratum appears in Br J Psychiatry 1998 Jul;173:89]. Br J Psychiatry. 1998;172: 485-90. PMID: 9828987.

76. Bazelmans E, Prins JB, Lulofs R, et al. Cognitive behaviour group therapy for chronic fatigue syndrome: a non-randomised waiting list controlled study. Psychother Psychosom. 2005;74(4): 218-24. PMID: 15947511.

77. Burgess M, Andiappan M, Chalder T. Cognitive behaviour therapy for chronic fatigue syndrome in adults: face to face versus telephone treatment: a randomized controlled trial. Behav Cogn Psychother. 2012;40(2): 175-91. PMID: 21929831.

78. Deale A, Chalder T, Marks I, et al. Cognitive behavior therapy for chronic fatigue syndrome: a randomized controlled trial. Am J Psychiatry. 1997;154(3): 408-14. PMID: 9054791.

79. Deale A, Husain K, Chalder T, et al. Long-term outcome of cognitive behavior therapy versus relaxation therapy for chronic fatigue syndrome: a 5-year follow-up study. Am J Psychiatry. 2001;158(12): 2038-42. PMID: 11729022.

80. Goudsmit EM, Ho-Yen DO, Dancey CP. Learning to cope with chronic illness. Efficacy of a multi-component treatment for people with chronic fatigue syndrome. Patient Educ Couns. 2009;77(2): 231-6. PMID: 19576714.

81. Hlavaty LE, Brown MM, Jason LA. The effect of homework compliance on treatment outcomes for participants with myalgic encephalomyelitis/chronic fatigue syndrome. Rehabil Psychol. 2011;56(3): 212-8. PMID: 21767035.

82. Jason L, Benton M, Torres-Harding S, et al. The impact of energy modulation on physical functioning and fatigue severity among patients with ME/CFS. Patient Educ Couns. 2009;77(2): 237-41. PMID: 19356884.

83. Jason LA, Roesner N, Porter N, et al. Provision of social support to individuals with chronic fatigue syndrome. J Clin Psychol. 2010;66(3): 249-58. PMID: 19902489.

84. Jason LA, Torres-Harding S, Friedberg F, et al. Non-pharmacologic interventions for CFS: a randomized trial. J Clin Psychol Med Settings. 2007;14(4): 275-96.

85. Knoop H, van der Meer JWM, Bleijenberg G. Guided self-instructions for people with chronic fatigue syndrome: randomised controlled trial. Br J Psychiatry. 2008;193(4): 340-1. PMID: 18827302.

86. Lopez C, Antoni M, Penedo F, et al. A pilot study of cognitive behavioral stress management effects on stress, quality of life, and symptoms in persons with chronic fatigue syndrome. J Psychosom Res. 2011;70(4): 328-34. PMID: 21414452.

87. Núñez M, Fernandez-Sola J, Núñez E, et al. Health-related quality of life in patients with chronic fatigue syndrome: group cognitive behavioural therapy and graded exercise versus usual treatment. A randomised controlled trial with 1year of follow-up. Clin Rheumatol. 2011;30(3): 381-9. PMID: 21234629.

88. O'Dowd H, Gladwell P, Rogers CA, et al. Cognitive behavioural therapy in chronic fatigue syndrome: a randomised controlled trial of an outpatient group programme. Health Technol Assess. 2006;10(37): iii-iv, ix-x, 1-121. PMID: 17014748.

89. Prins JB, Bleijenberg G, Bazelmans E, et al. Cognitive behaviour therapy for chronic fatigue syndrome: a multicentre randomised controlled trial. Lancet. 2001;357(9259): 841-7. PMID: 11265953.

90. Sharpe M, Hawton K, Simkin S, et al. Cognitive behaviour therapy for the chronic fatigue syndrome: a randomized controlled trial. BMJ. 1996;312(7022): 22-6. PMID: 8555852.

91. Taylor RR. Quality of life and symptom severity for individuals with chronic fatigue syndrome: findings from a randomized clinical trial. Am J Occup Ther. 2004;58(1): 35-43. PMID: 14763634.

92. Tummers M, Knoop H, Bleijenberg G. Effectiveness of stepped care for chronic fatigue syndrome: a randomized noninferiority trial. J Consult Clin Psychol. 2010;78(5): 724-31. PMID: 20873907.

93. Tummers M, Knoop H, van Dam A, et al. Implementing a minimal intervention for chronic fatigue syndrome in a mental health centre: a randomized controlled trial. Psychol Med. 2012;42(10): 2205-15. PMID: 22354999.

94. Tummers M, Knoop H, van Dam A, et al. Moderators of the treatment response to guided self-instruction for chronic fatigue syndrome. J Psychosom Res. 2013;74(5): 373-7. PMID: 23597323.

95. Wearden AJ, Dowrick C, Chew-Graham C, et al. Nurse led, home based self help treatment for patients in primary care with chronic fatigue syndrome: randomised controlled trial. BMJ. 2010;340: c1777. PMID: 20418251.

96. Wearden AJ, Dunn G, Dowrick C, et al. Depressive symptoms and pragmatic rehabilitation for chronic fatigue syndrome. Br J Psychiatry. 2012;201(3): 227-32. PMID: 22844025.

97. Wearden AJ, Emsley R. Mediators of the effects on fatigue of pragmatic rehabilitation for chronic fatigue syndrome. J Consult Clin Psychol. 2013;81(5): 831-8. PMID: 23796316.

98. White PD, Goldsmith KA, Johnson AL, et al. Comparison of adaptive pacing therapy, cognitive behaviour therapy, graded exercise therapy, and specialist medical care for chronic fatigue syndrome (PACE): a randomised trial. Lancet. 2011;377(9768): 823-36. PMID: 21334061.

99. Hobday RA, Thomas S, O'Donovan A, et al. Dietary intervention in chronic fatigue syndrome. J Hum Nutr Diet. 2008;21(2): 141-9. PMID: 18339054.

100. Öckerman PA. Antioxidant treatment of chronic fatigue syndrome. Clin Pract Alternat Med. 2000;1(2): 88-91.

101. The GKH, Bleijenberg G, van der Meer JWM. The effect of acclydine in chronic fatigue syndrome: A randomized controlled trial. PLoS Clinical Trials. 2007;2(5)PMID: 17525791.

102. Vermeulen RCW, Scholte HR. Exploratory open label, randomized study of acetyl- and propionylcarnitine in chronic fatigue syndrome. Psychosom Med. 2004;66(2): 276-82. PMID: 15039515.

103. Walach H, Bosch H, Lewith G, et al. Effectiveness of distant healing for patients with chronic fatigue syndrome: a randomised controlled partially blinded trial (EUHEALS). Psychother Psychosom. 2008;77(3): 158-66. PMID: 18277062.

104. Weatherley-Jones E, Nicholl JP, Thomas KJ, et al. A randomised, controlled, triple-blind trial of the efficacy of homeopathic treatment for chronic fatigue syndrome. J Psychosom Res. 2004;56(2): 189-97. PMID: 15016577.

105. Williams G, Waterhouse J, Mugarza J, et al. Therapy of circadian rhythm disorders in chronic fatigue syndrome: no symptomatic improvement with melatonin or phototherapy. Eur J Clin Invest. 2002;32(11): 831-7. PMID: 12423324.

106. Fulcher KY, White PD. Randomised controlled trial of graded exercise in patients with the chronic fatigue syndrome. BMJ. 1997;314(7095): 1647-52. PMID: 9180065.

107. Ho RTH, Chan JSM, Wang C-W, et al. A randomized controlled trial of qigong exercise on fatigue symptoms, functioning, and telomerase activity in persons with chronic fatigue or chronic fatigue syndrome. Ann Behav Med. 2012;44(2): 160-70. PMID: 22736201.

108. Moss-Morris R, Sharon C, Tobin R, et al. A randomized controlled graded exercise trial for chronic fatigue syndrome: outcomes and mechanisms of change. J Health Psychol. 2005;10(2): 245-59. PMID: 15723894.

109. Sutcliffe K, Gray J, Tan MP, et al. Home orthostatic training in chronic fatigue syndrome--a randomized, placebo-controlled feasibility study. Eur J Clin Invest. 2010;40(1): 18-24. PMID: 19912315.

110. Vanness JM, Snell CR, Strayer DR, et al. Subclassifying chronic fatigue syndrome through exercise testing. Med Sci Sports Exerc. 2003;35(6): 908-13. PMID: 12783037.

111. Sisto SA, LaManca J, Cordero DL, et al. Metabolic and cardiovascular effects of a progressive exercise test in patients with chronic fatigue syndrome. Am J Med. 1996;100(6): 634-40. PMID: 8678084.

112. Biswal B, Kunwar P, Natelson BH. Cerebral blood flow is reduced in chronic fatigue syndrome as assessed by arterial spin labeling. J Neurol Sci. 2011;301(1-2): 9-11. PMID: 21167506.

113. Boda WL, Natelson BH, Sisto SA, et al. Gait abnormalities in chronic fatigue syndrome. J Neurol Sci. 1995;131(2): 156-61. PMID: 7595641.

114. Frith J, Zalewski P, Klawe JJ, et al. Impaired blood pressure variability in chronic fatigue syndrome--a potential biomarker. QJM. 2012;105(9): 831-8. PMID: 22670061.

115. Naschitz JE, Sabo E, Naschitz S, et al. Hemodynamics instability score in chronic fatigue syndrome and in non-chronic fatigue syndrome.[Erratum appears in Semin Arthritis Rheum. 2003 Apr;32(5):343 Note: Madelain, Fields [corrected to Fields, Madeline]; Hillel, Isseroff [corrected to Isseroff, Hillel]]. Semin Arthritis Rheum. 2002;32(3): 141-8. PMID: 12528078.

116. Jones DEJ, Hollingsworth KG, Jakovljevic DG, et al. Loss of capacity to recover from acidosis on repeat exercise in chronic fatigue syndrome: a case-control study. Eur J Clin Invest. 2012;42(2): 186-94. PMID: 21749371.

117. Scadding JG. Diagnosis: the clinician and the computer. Lancet. 1967;2(7521): 877-82. PMID: 4168324.

118. Brurberg KG, Fonhus MS, Larun L, et al. Case definitions for chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME): a systematic review. BMJ Open. 2014;4(2): e003973. PMID: 24508851.

119. Huibers MJ, Wessely S. The act of diagnosis: pros and cons of labelling chronic fatigue syndrome. Psychol Med. 2006;36(7): 895-900. PMID: 16403245.

120. Komaroff AL, Buchwald DS. Chronic fatigue syndrome: an update. Annu Rev Med. 1998;49: 1-13. PMID: 9509246.

121. Komaroff AL. Is human herpesvirus-6 a trigger for chronic fatigue syndrome? J Clin Virol. 2006;37 Suppl 1: S39-46. PMID: 17276367.

122. Devanur LD, Kerr JR. Chronic fatigue syndrome. J Clin Virol. 2006;37(3): 139-50. PMID: 16978917.

123. Straus SE, Dale JK, Tobi M, et al. Acyclovir treatment of the chronic fatigue syndrome. Lack of efficacy in a placebo-controlled trial. N Engl J Med. 1988;319(26): 1692-8. PMID: 2849717.

124. Lloyd A, Hickie I, Wakefield D, et al. A double-blind, placebo-controlled trial of intravenous immunoglobulin therapy in patients with chronic fatigue syndrome. Am J Med. 1990;89(5): 561-8. PMID: 2146875.

125. Chambers D, Bagnall A-M, Hempel S, et al. Interventions for the treatment, management and rehabilitation of patients with chronic fatigue syndrome/myalgic encephalomyelitis: an updated systematic review. J R Soc Med. 2006;99(10): 506-20. PMID: 17021301.

126. Whiting P, Bagnall AM, Sowden AJ, et al. Interventions for the treatment and management of chronic fatigue syndrome: a systematic review. JAMA. 2001;286(11): 1360-8. PMID: 11560542.

127. Alraek T, Lee MS, Choi T-Y, et al. Complementary and alternative medicine for patients with chronic fatigue syndrome: a systematic review. BMC Altern Med. 2011;11: 87. PMID: 21982120.

128. Kim J-E, Hong K-E, Kim H-J, et al. An open-label study of effects of acupuncture on chronic fatigue syndrome and idiopathic chronic fatigue: study protocol for a randomized controlled trial. Trials. 2013;14: 147. PMID: 23693129.

129. Cho HJ, Hotopf M, Wessely S. The placebo response in the treatment of chronic fatigue syndrome: a systematic review and meta-analysis. Psychosom Med. 2005;67(2): 301-13. PMID: 15784798.

130. Cooper L. Report on Survey of Members of Local ME Groups. Bristol, UK: Action for ME. 2000. Availible at: http://www.actionforme.org.uk/Resources/Action%20for%20ME/Documents/get-informed/Local%20Group%20Survey%202001%20Lesley%20Cooper.PDF. Accessed July 30, 2014.

131. Edmonds M, McGuire H, Price J. Exercise therapy for chronic fatigue syndrome. Cochrane Database Syst Rev. 2004(3): CD003200. PMID: 15266475.

Abbreviations and Acronyms

ACTH	adrenocorticotropic hormone
AHRQ	Agency for Healthcare Research and Quality
AUC	area under the curve
AUROC	area under the receiver operating characteristic
CAM	complementary and alternative medicine
CBT	cognitive behavioral therapies
CDC	Centers for Disease Control and Prevention
CFS	chronic fatigue syndrome
CGI	Clinical Global Impression Change
CI	confidence interval
CIS	Checklist of Individual Strength
CPET	Cardiopulmonary exercise test
EPC	Evidence-based Practice Center
FDA	Food and Drug Administration
FIQ	Fibromyalgia Impact Questionnaire
FIS	Fatigue Impact Scale
FLP	Functional Limitations Profile
FSS	Fatigue Severity Scale
GET	graded exercise treatment
HADS	Hospital Anxiety and Depression Scale
HADS-A	anxiety subscale of HADS
HADS-D	depression subscale of HADS
HR	hazard ratio
IGF-1	insulin-like growth factor one
KPS	Karnofsky Performance Scale
ME	myalgic encephalomyelitis
MFI-20	Multidimensional Fatigue Inventory, 20-item
MOS-SF	Medical Outcome Study Short Form
NIH	National Institutes of Health
ODP	Office of Disease Prevention
OR	odds ratio
PEM	post exertional malaise
PICOTS	populations, interventions, comparators, outcomes, timing, setting
POMS	Profile of Mood States
QLI	Quality of Life Index
QLS	Quality of Life Scale
QOLI	Quality of Life Inventory
RA	rheumatoid arthritis
ROC	receiver operating curve
RR	relative risk
SCL-90-R	Symptom Checklist-90-revised
SD	standard deviation
SF-12	Short Form 12-item Health Survey

SF-36	36-item Short Form Survey
SIP-8	Sickness Impact Profile 8-item
SOFA-CFS	Schedule of Fatigue and Angina for CFS scale
TEP	technical expert panel
TSST	Trier Social Stress Test
VAS	visual analogue scale